

Formalių konceptų analizė, kuriant taikomosios srities ontologijas

Darius Jurkevičius

Vilniaus Gedimino technikos universiteto
Informacinių sistemų katedros doktorantas
Vilnius Gediminas Technical University,
Information Systems Department,
Doctoral student
Saulėtekio al. 11, LT-10223 Vilnius
El. paštas: d.jurkevicius@isl.vgtu.lt

Algirdas Laukaitis

Vilniaus Gedimino technikos universiteto
Informacinių sistemų mokslo laboratorijos
docentas, daktaras
Vilnius Gediminas Technical University,
Information Systems Research Laboratory,
Associate Professor, Doctor
Saulėtekio al. 11, LT-10223 Vilnius
El. paštas: algirdas.laukaitis@fm.vgtu.lt

Olegas Vasilecas

Vilniaus Gedimino technikos universiteto
Informacinių sistemų mokslo laboratorijos
profesorius, daktaras
Vilnius Gediminas Technical University,
Information Systems Research Laboratory
Professor, Doctor
Saulėtekio al. 11, LT-10223 Vilnius
Tel. (+370 5) 274 48 60
El. paštas: Olegas.Vasilecas@fm.vgtu.lt

Straipsnyje pristatomas naujas metodas, kuris padeda sukurti dalykinės srities ontologiją. Pagrindinis metodo bruožas pasireiškia tuo, kad ontologijos kūrimo metu dalykinės srities objektai, konceptai, bei konceptų atributai analizuojami atliekant formalių konceptų analizės algoritmus. Be to, šio metodo ontologijos kūrimo rezultatai gali būti taikomi dalykinės srities dokumentų semantiniam indeksavimui, kas ypač svarbu plėtojant semantinio interneto technologijas. Be šio metodo pristatymo, šiame straipsnyje yra aprašytas eksperimentinis tyrimas susijęs su nekilnojamojo turto dalykine sritimi. Eksperimento metu buvo sukurtas sistemos prototipas, kuris realizuoja straipsnyje pateiktą ontologijos kūrimo metodą. Manome, kad sukurto prototipo architektūra taip pat bus naudinga kitiems tyrėjams ontologijų kūrimo srityje.

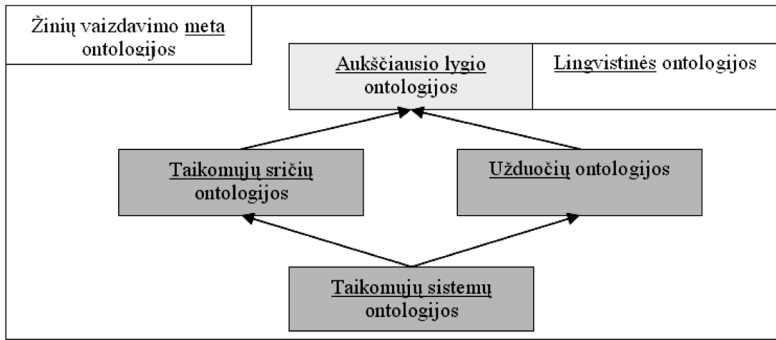
Pagrindiniai žodžiai: ontologija, formalių konceptų analizė, formalus konceptas, duomenų analizė.

1. Įvadas¹

Informacinių sistemų inžinerijoje paprastai naudojamos skirtingų tipų (1 pav.) hierarchinės ontologijos. Sudarant ontologiją viršutiniu lygmeniu apibrėžiamos bendriausios

sąvokos (esybė, įvykis, laikas, procesas ir kt.). Kitu lygmeniu aprašomos dalykinės srities sąvokos (dalykinės srities ontologija), vykstančių procesų aprašymo sąvokos (proceso ontologija) ir sprendžiamų uždavinių aprašymo sąvokos (problemos ontologija). Šitaip įgyvendinamas N. Guarino pasiūlytas (Guarino, 1997) žinių apie dalykinę sritį ir žinių apie procesus nepriklausomumo principas. Procesai aprašomi vaidmenų terminais ir susiejami su problema priskiriant

¹ Šis darbas yra remiamas Lietuvos valstybinio mokslo ir studijų fondo pagal aukštųjų technologijų kūrimo programos projektą „Veiklos taisyklių sprendimai informacinių sistemų kūrimui (VeTIS)“, registracijos nr. B-07042.



1 pav. Ontologijų tipai

tiems vaidmenims vienas ar kitas dalykinės srities esybes. Ontologijos, kaip ir kiti informacinių sistemų komponentai, gali būti daugkartinio naudojimo, nes tos pačios sąvokos vartojamos skirtingoms dalykinėms sritims konceptualizuoti. Daugkartinio taikymo yra ir procesų ontologijos, nes skirtingose dalykinėse srityse vykstantys procesai dažnai turi tą pačią struktūrą.

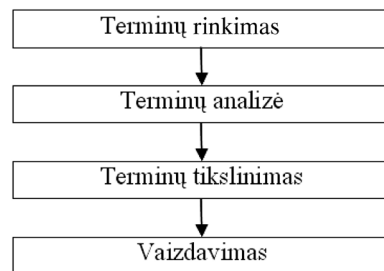
Paprastai ontologijos kūrimo procesą sudaro keturi etapai: terminų rinkimas, jų analizė, tikslinimas ir vaizdavimas. Šio proceso schema pavaizduota 2 paveiksle.

Terminų surinkimo etape yra identifikuojami visi nagrinėjamos srities vartojami terminai, jų tarpusavio ryšiai ir apibrėžimai. Analizės etape surinkti terminai semantiškai analizuojami: ieškoma skirtingų terminų, apibūdinančių vienodus objektus ar reiškinius. Radus šiuos terminus, turi būti pasiektas susitarimas dėl vieno termino bendro vartojimo. Tai daroma trečiajame ontologijos sudarymo etape. Peržiūrėjus visus tokius terminus pereinama prie paskutinio etapo – ontologijos vaizdavimo pasirinkta ontologijų vaizdavimo kalba. Paskutinis etapas gali būti įgyvendinamas taikant pasirinktą ontologijų kūrimo priemonę. Tai lemia ir ontologijos vaizdavimo kalbos

pasirinkimą. Taikant ontologijų kūrimo priemones ontologijas gali kurti žemesnės kvalifikacijos darbuotojai, pavyzdžiui, imonės informacinių sistemų specialistai.

Šis anksčiau aprašytas ontologijos kūrimo procesas yra gana lėtas ir čia reikia didelio kruopštumo.

Vienas iš būdų pagreitinti ontologijos kūrimą yra terminų analizės ir ontologijos vaizdavimo etapų automatizavimas. Šiuo metu nėra sukurta sistemų, leidžiančių automatizuoti minėtus ontologijos kūrimo etapus. Tačiau ontologijų kūrimo srities tyrėjų bendruomenė supranta, kad, norint semantinio interneto technologijų proveržio, reikia sistemų, kurios bent iš dalies automatizuotų ontologijos kūrimo etapus. Šiame straipsnyje aprašomas metodas pateikia būdus, kaip galima automatizuoti



2 pav. Ontologijos kūrimo procesas

ontologijos kūrimo procesus taikant formalios konceptų analizės metodus.

Likusi straipsnio dalis susideda iš dalių: 2 skyriuje aprašomi formalių konceptų analizės principai bei jų panaudojimas kuriant ontologijas; 3 skyriuje pateikiami dalykinės srities terminų surinkimo principai, kuriuos naudojame ontologijos projektavimo metu; 4 skyriuje aprašomas eksperimentinis tyrimas bei eksperimento metu sukurtos ontologijų projektavimo sistemos prototipo architektūra. Straipsnis baigiamas rezultatų aptarimu, pateikiama tolesnių tyrimų analizė.

2. Formalūs konceptai ir jų vartojimas ontologijoms kurti

Formalių konceptų analizė yra metodas, taikomas duomenų analizei, žinių vaizdavimui ir informacijos valdymui. Formalių konceptų analizė buvo pasiūlyta Rudolfo Wille 1981 metais (Wille, 1982). Pirmuosius 10 metų FKA tyrinėjo mažos grupelės mokslininkų ir Rudolfo Wille studentai Vokietijoje. FKA nebuvo plačiai žinoma dėl to, kad dauguma publikacijų buvo skelbiama matematikų konferencijose. Finansuojant tyrimus šioje srityje buvo įgyvendinti keli dideli projektai, kurių didžioji dalis buvo žinių tyrimo sistema, naudojama statybose. Ši sistema nebuvo žinoma už Vokietijos ribų. Tačiau per paskutinius 10 metų FKA tapo tarptautinės mokslininkų bendruomenės tyrimo objektu. FKA buvo pritaikyta lingvistikoje, psichologijoje, taip pat programinei įrangai projektuoti, dirbtinio intelekto ir informacijos paieškos srityse.

Kai kurie iš FKA struktūrų informacijai pateikti buvo suformuluoti atskirų nepriklausomų tyrinėtojų. Pavyzdžiui, Godin ir

kiti naudojo konceptų tinklus (angl. *Galois lattices*) informacijos paieškai.

Siekdami pristatyti sukurtą metodo FKA panaudojimo principus apibrėšim FKA vartojamus terminus:

Koncepto apibrėžimas susideda iš tokių dalių:

- abstrakti idėja, kuri apibendrina atskirus egzempliorius (objektus);
- turi intenciją (nuodugną apibrėžimą), ekstensiją (objektų arba egzempliorių aibę) (Martin, 1994);
- turi intensionalą (nustato narystės kriterijus), ir ekstensionalą (aibė objektų arba egzempliorių).

Formalus konceptas apibrėžiamas pagal (Wormuth, 2004):

- jį sudaro aibė objektų, kurie turi aibę bendrų atributų. Ši sąvoka taikoma formalių konceptų analizėje.

Formalių konceptų analizė (FKA) (Wormuth, 2004) yra metodas, skirtas:

- matematiškai aprašyti filosofines sąvokas;
- žmonėms pritaikytas metodas duomenims analizuoti ir juos vaizduoti struktūriniu būdu;
- vaizduoti duomenis, jų struktūras, reikšmes ir priklausomybes.

Formalus kontekstas:

- objektų, atributų ir ryšių tarp objektų ir atributų visuma vadinama formaliu kontekstu (G, M, I), kur G – objektai, M – atributai, o I – objektų ir atributų ryšys. Ryšys žymimas šitaip: jeigu objektas turi atributą, tai susikirtimas žymimas 1. Reikia pažymėti, kad Dau (Dau, 2004) pastebėjo, kad mokslininkai, braižydami diagramas, negalėjo jais remtis kaip argumentais.

Formaliai atskiriant matematinę struktūrą nuo jos schematinio vaizdavimo buvo sudaryta darbo aplinka, kurioje diagramas galima daryti formaliems pagrindimams.

Dabar pateiksime formalių konceptų analizės aprašymą ir apibrėžimus (Ganter, 1999). Padarykime prielaidą, kad G yra objektų aibė, kuri priklauso tam tikrai dalykinei sričiai (pavyzdžiui: namas, butas, pirkėjas ir t. t.). Tarkim, M yra atributų aibė, kuri nusako mūsų nagrinėjamų objektų charakteristikas. Mes identifikuojame indeksą I kaip binarinį sąryšį tarp aibių G ir M , t. y. $I \subseteq GPM$. Užrašas (G, M, I) vadinamas formalių kontekstu. Apibrėžiam $A \subseteq G$, kaip

$$A' := \{m \in M \mid (g, m) \in I \text{ for all } g \in A\} \quad (1)$$

ir kitą dalį, $B \subseteq M$, kaip

$$B' := \{m \in G \mid (g, m) \in I \text{ for all } g \in B\} \quad (2)$$

Formalaus konteksto (G, M, I) formalus konceptas apibrėžiamas kaip pora (A, B) , kai $A \subseteq G$, $B \subseteq M$, $A' \subseteq B$ ir $B' \subseteq A$. Kitaip tariant, formalus konceptas – tai pora (A, B) , kur $A' = B$, $B' = A$. Formalaus konteksto (G, M, I) formalių konceptų aibė, vadinama konteksto (G, M, I) koncepciniu tinklu.

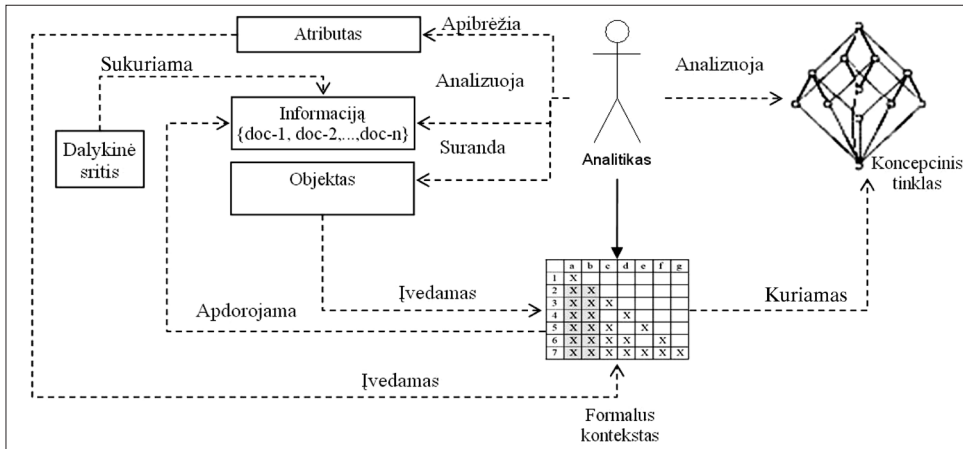
Kuriant ontologiją, dalykinės srities surinktiems terminams saugoti siūlome naudoti formalų kontekstą, kad vėliau terminus būtų galima analizuoti, tikslinti bei vaizduoti. Taip išsaugotus formaliame kontekste terminus galima vartoti įvairioms užduotims atlikti. Pavyzdžiui, paieškai, informacijos apdorojimui ir t. t.

Grafiškai atvaizdavus ontologiją ir ją analizuojant galima surasti duomenų priklausomybes, atitikimus, pasikartojančias struktūras, išimtis. Žinių arba ontologijų

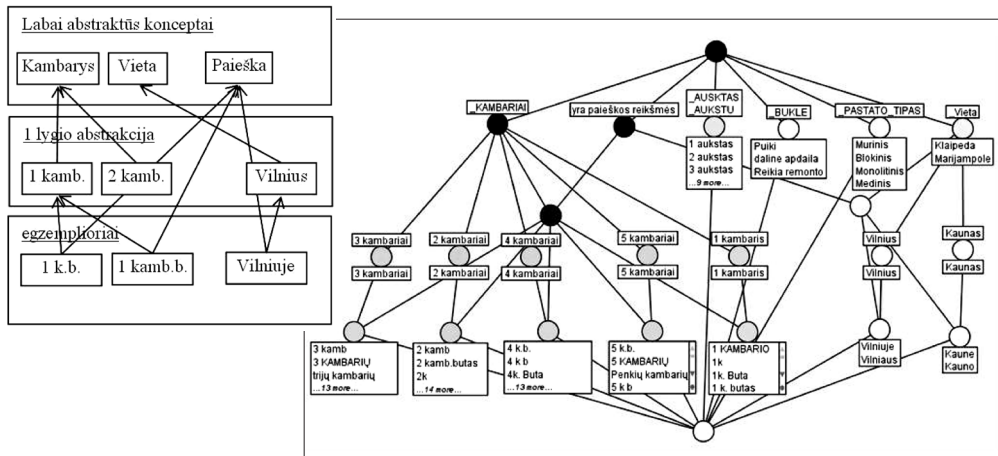
vaizdavimas yra svarbi intelektualios duomenų analizės dalis. Daug žmonių geriau supranta vaizdinę informaciją, kuri leidžia greičiau surasti priklausomybes, kurias įprastais metodais galima būtų rasti tik pasitelkus daugelio skaičiavimų duomenų analizę. Grafiškai ontologijos vaizduojamos medžių, arba hierarchine forma.

3. Dalykinės srities terminų surinkimas ir duomenų analizė

Šiame skyriuje aprašysime mūsų siūlomą terminų surinkimo ir duomenų analizės metodą (3 pav.). 3 pav. pavaizduotas analitiskas pirmiausia apibrėžia labai abstrakčius konceptus (sukuriami objektai ir atributai tuo pačiu pavadinimu) bei atributus, kurie vartojami užduotims (užduotis aprašo, kur bus vartojami surinkti terminai) sudaryti (3 pav.). Būsima ontologija kuriama rinkinių užduotims (procesų) atlikti. Kaip pavyzdį, ontologiją taikysime paieškai atlikti (užduotis – paieška). 4 paveiksle yra parodytas užduoties atributas pavadinimu „Paieška“ bei labai abstraktūs konceptai (objektai) „kambarys“ ir „vieta“. Toliau įrašomi konceptai (objektai) (1 lygio abstrakcija), kurie priklauso labai abstraktiems konceptams (objektams). Taip formuojamas formalių konceptų hierarchinis medis. Šie konceptai pridedami pagal poreikį analizuojant turimus duomenis (3, 5 pav.). Duomenų analizės yra sukuriamas specialus šablonas (5 pav.), kur yra pateikiami turimi susiję su dalykine sritimi duomenys, rinkinys atributų, kurie yra susiję su užduotimis (procesais). Vartojant surastus analizės metu konceptus (objektus), galima juos vartoti atliekant informacijos analizę ieškant kitų terminų (objektus) (5 pav.). Yra terminų



3 pav. Dalykinės srities terminų surinkimas, analizė ir vaizdavimas



4 pav. Kairė pusė: pasiskirstymas pagal abstrakcijos lygį. Dešinė pusė: koncepcinis tinklas, sukurtas naudojant ToscanaJ-1.6 programinį įrankį

(objektų) ryšys su užduoties atributu (pvz., Paeška), tam, kad vėliau juos taikant galima būtų atskirti, koks terminas vartojamas, kokiai užduočiai. Atliekant paiešką turimoje informacijoje ir, suradus tam tikrą terminą, jis priskiriamas tam tikram aukštesnio lygio konceptui ir įtraukiamas į formalų kontekstą. Atliekant duomenų analizę ir pildant surastais objektais formalų kontekstą, natūraliai yra sudaroma dalykinės srities konceptų hierarchinis medis (ontologija).

Suradus visus reikiamus konceptus bei terminus formaliam kontekstui, šiuos duomenis galima analizuoti sukūrus, taikant formalų konceptų analizės metodą, koncepcinį tinklą (4 pav.). Remdamasis sukurtu konceptų medžiu, analitikas gali surasti priklausomybes, atitikimus, pasikartojančias struktūras, išimtis ir t. t. tam, kad vėliau jis galėtų sudaryti dalykinės srities ontologiją vartodamas įprastą ontologijų vaizdavimo kalbas.

NEKILNOJAMOJO TURTO PAIEŠKOS SISTEMA

Paieška Ontologija

Eil. Nr.	SKELBIMAS	ANTRAŠTE	NT TIPAS	VEIKSMAS	KAMBARIAI	BŪKLĖ	VIETA	DATA
1	Ieskau nuomai kambariobuto Vilniuje nuo birželio 1d.	Išsinuomosisiu kambari/buta Vilniuje nuo birželio 1d.	butas	Išnuomoti	INSERT	INSERT	Vilnius	2007-05-16
2	Parduodu Naujamiestyje, Zemaiteis g.71, gerame apituoktame mur name, (toku namu Vilniuje nedaug.) 2 kam b. 51 kv.m., 2/5, kambariai izoliuoti, didele lodzija per visa siena langas, langai liseina i pietine ir rytine puse, kiema Butas labai sviesus, jaukus ir	Parduodu Naujamiestyje, Zemaiteis g. mur. n. 2km. b.	butas butas	Parduoti Parduoti Parduoti	2 kambariai	INSERT	Vilnius	2007-05-15
3	Išnuomosisiu 1 k. buta Vaduvos g. Vilniuje su baldais	Išnuomosisiu 1 k. buta Vaduvos g.	butas butas	Išnuomoti Išnuomoti	1_kambarys 1_kambarys	INSERT	Vilnius	2007-05-14
4	Nuomoji nauja 2k buta netoli Vilniaus centro, yra garazas, nauji baldai ir buities technika, moderniai rengtas, rami vieta, galima ivairiam laikotarpiui, kaina sutartine, tel: 867140452; 868306953	Nuomoji nauja 2k buta netoli Vilniaus centro	butas butas	Išnuomoti Išnuomoti	2_kambariai 2_kambariai	INSERT	Vilnius	2007-05-12
5	Tvarkinga dirbanti pora išsinuomos 1 kambario butą Vilniuje ilgesniam laikui. Mokesis iki 600 lt.	Išsinuomosisiu 1 kambario butą	butas butas	INSERT	1_kambarys 1_kambarys	tvarkingas	Vilnius	2007-05-11
6	Jauna tvarkinga pora išsinuomotų 1 kambario butą Vilniuje. Ilgesniam laikui Mokesis iki 700lt.	Iškom išsinuomoti butą ilgesniam laikui	butas butas	Išnuomoti Pirkti	1_kambarys 1_kambarys	tvarkingas	Vilnius	2007-05-10

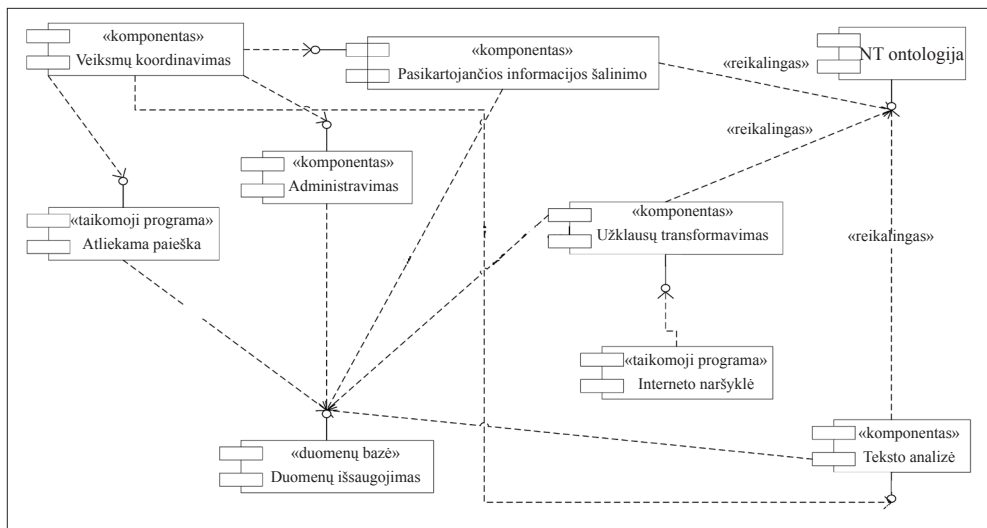
5 pav. Terminų suradimas ir įrašymas į formalų kontekstą

4. Eksperimentinis tyrimas

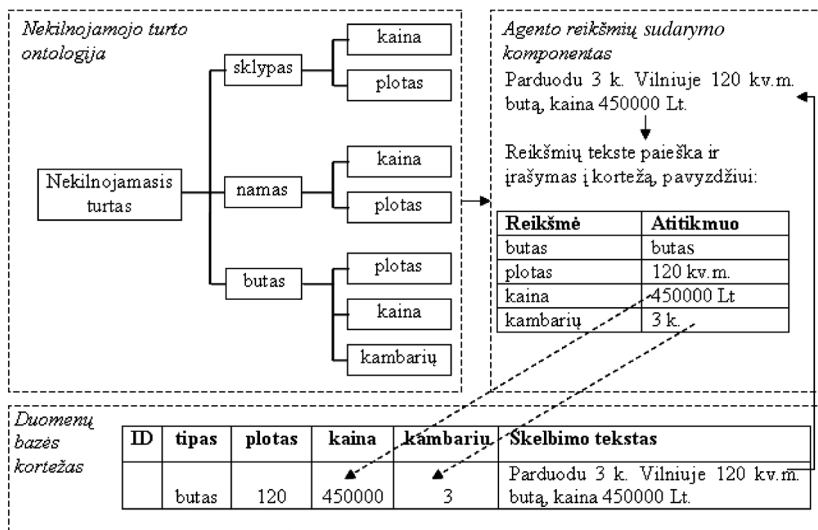
Eksperimentui buvo naudojama paieškos, saugojimo ir analizės programa (6 pav.), kurios tikslas yra surinkti iš interneto tinklapių duomenis ir juos apdoroti taikant dalykinės srities ontologiją bei pateikti apibendrintą informaciją vartotojui.

Kuriant ontologiją buvo išnagrinėta ir išanalizuota nekilnojamojo turto skelbimų

aibė. Šie skelbimai buvo surinkti iš šių pagrindinių interneto svetainių, kuriose jie yra skelbiami, tai duomenys apie nekilnojamojo turto sandorius bei pasiūlymus. Kitas sąrašas pateikia nuorodas į svetaines, iš kurių buvo surinkti skelbimai: <http://www.aruodas.lt>, <http://www.skelbiu.lt>, <http://www.domoplius.lt>, <http://www.alioreklama.lt>, <http://www.skelbimai.lt>, <http://www.edomus.lt>, <http://www.



6 pav. Informacijos paieškos, saugojimo ir analizės architektūra



7 pav. Informacijos apdorojimo proceso schema

enamai.lt>, <http://www.city24.lt>, <http://www.namai.lt>, <http://www.muge.lt>.

Medžiagai, susijusiai su anksčiau minėta dalykine sritimi, surinkti ir išsaugoti buvo naudojamas WebScrafer Plus + programinis įrankis, kuris juos surinko ir išsaugojo duomenų bazėje.

Siūlomo metodo kitas žingsnis yra atliekamas informacijos apdorojimas, kurio metu informacijoje yra surandami atitikimai pagal įrašytus terminus (objektus) (7 pav.). Šiai užduočiai atlikti reikalinga nekilnojamojo turto ontologija.

Todėl toliau aprašysime nekilnojamojo turto ontologijos kūrimą. Buvo sukurtos dvi

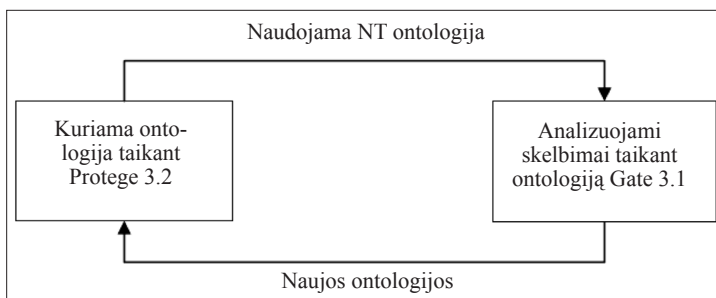
ontologijos, viena įprastais metodais taikant Protege programinį įrankį, o kita taikant šiame straipsnyje pasiūlytą metodą.

4.1. Ontologijos kūrimas taikant įprastą metodą

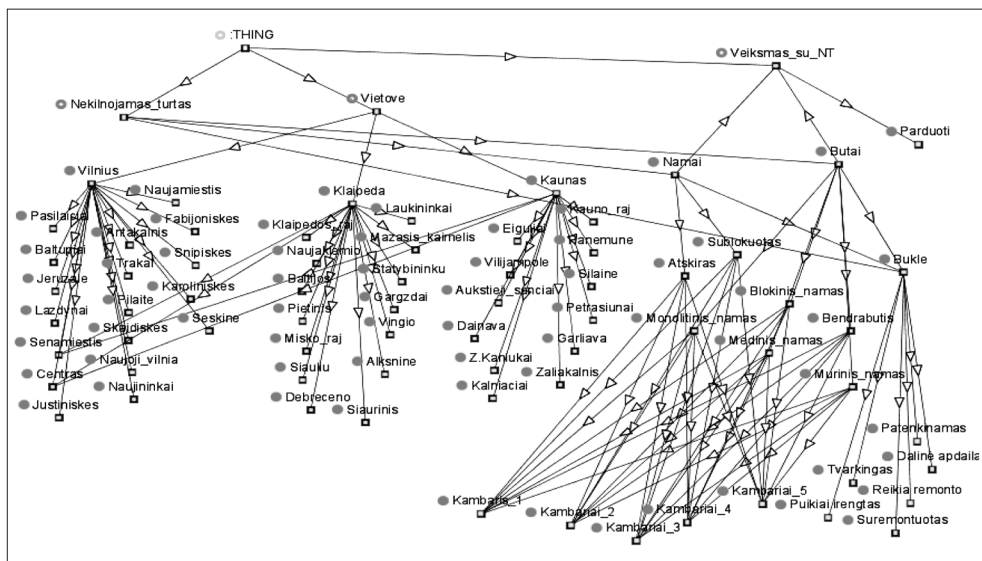
Atliekant eksperimentą buvo remiamasi įprastu (1 pav.) ontologijos kūrimo procesu.

Terminų analizei ir tikslinimui buvo taikomi Protege 3.2 bei Gate 3.1 nemokami programiniai įrankiai. Atliktas terminų analizės procesas parodytas 8 pav.

Gauta nekilnojamojo turto ontologija buvo vaizduojama taikant ontologijų kūri-



8 pav. Nekilnojamojo turto ontologijos kūrimo procesas taikant Protege ir Gate



9 pav. Nekilnojamojo turto ontologijos hierarchinė schema

mo įrankį Protege. Taigi gauta ontologija pavaizduota 9 paveiksle.

4.1.1. Klasių kūrimas

Atlikus surinktų duomenų analizę yra sukurtas klasių medis (9 pav.): Nekilnojamasis turtas, vietovė, veiksmas su NT. Kiekvienos klasės yra sukurti poklasiai.

- ◆ Klasė „nekilnojamasis turtas“: namai; butai; būklė.
- ◆ Klasė „veiksmas su NT“: parduoti; pirkti; išnuomoti; keisti.
- ◆ Klasė „Vietovė“ yra:
 - Vilnius:
 - Naujamiestis;
 - Fabijoniškės;
 - Šnipiškės;
 - Karoliniškės ir t. t.
 - Klaipėda:
 - Laukininkai;
 - Alksninė;
 - Pietinis ir t. t.

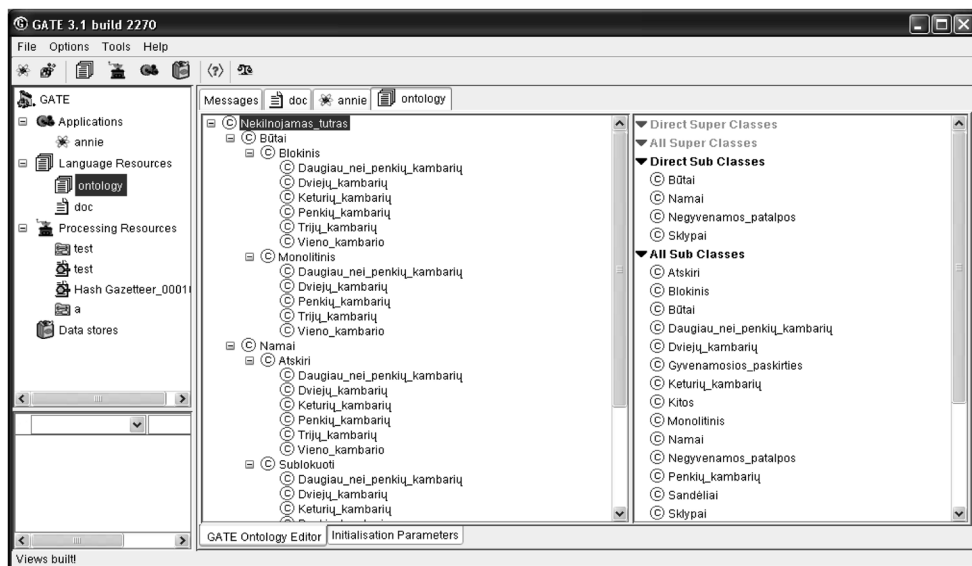
- Kaunas:
 - Šilainai;
 - Dainava;
 - Žaliakalnis ir t. t.

4.1.2. Ontologijos pildymas

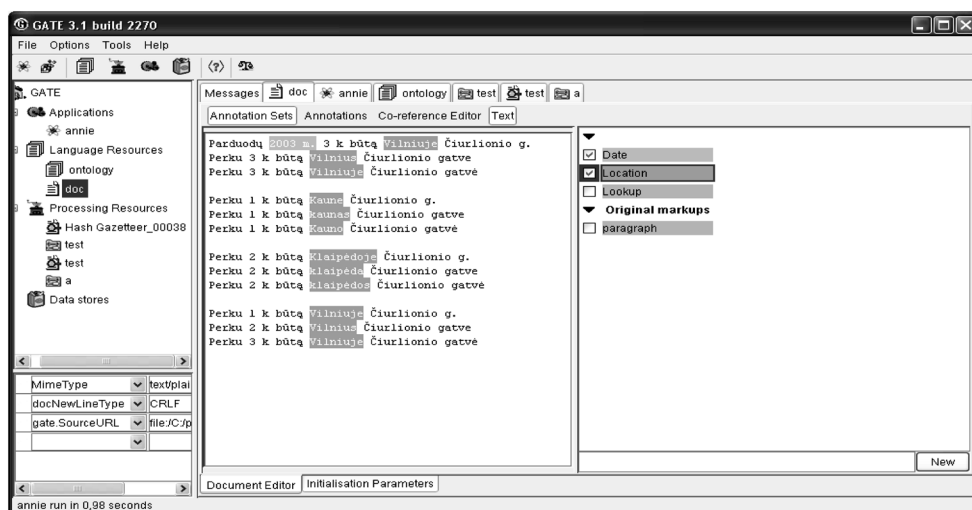
Darant surinktų skelbimų analizę buvo išskirtos reikiamos ontologijos, taikomas nemokamai platinamas Gate 3.1 programinis įrankis (10, 11 pav.). Naudojant Gate programinį įrankį buvo anotuojamas tekstas, išskiriami tam tikros rūšies terminai.

4.2. Ontologijos kūrimas taikant siūlomą metodą

Siekiant atlikti eksperimentą buvo sukurtas įrankis „Nekilnojamojo turto paieškos sistema“, kuris leidžia pildyti duomenų bazėje esantį formalų kontekstą naudojant naršyklę (5 pav.). Paieškos bei duomenų analizės formos yra sudaromos naudojant turimą formalų kontekstą. Tai užtikrina dinaminį formų pasikeitimą atsižvelgiant į įrašomus



10 pav. Kuriamas nekilnojamojo turto ontologija analizuojant skelbimus



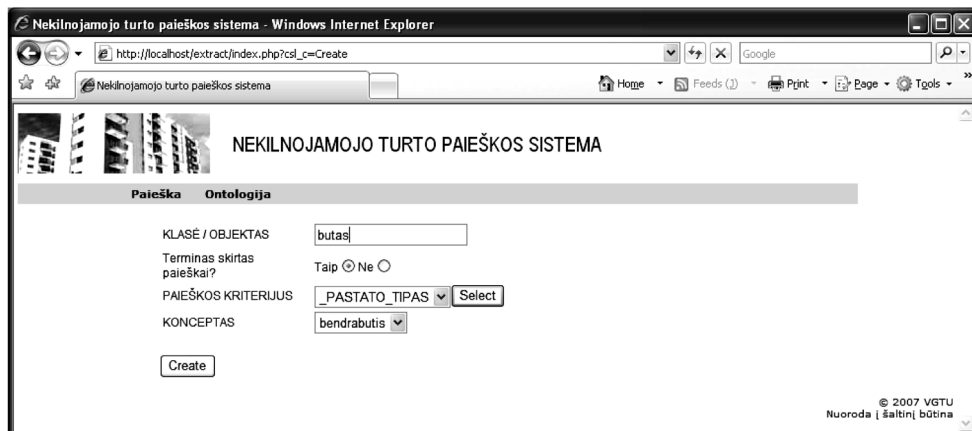
11 pav. GATE taikymas tekstui analizuoti

duomenis. Įrankis sukurtas vartojant PHP kalbą ir skirtas dirbti MySQL duomenų bazių valdymo sistema.

Siekiant eksperimentą atlikti, buvo naudojami 11465 skelbimai nekilnojamojo turto tematika. Buvo pasirinkta atributo „Paieška“ užduotis. Toliau pateiksime detalesnį aprašymą.

4.2.1. Formalaus konteksto pildymas

Sukurtas įrankis „Nekilnojamojo turto paieškos sistema“, kuris leidžia pildyti duomenų bazėje esantį formalų kontekstą naudojant žiniatinklio naršyklę (12, 14 pav.). Įrankis sukurtas vartojant PHP kalbą ir skirtas dirbti MySQL duomenų bazių



12 pav. Objekto įrašymas į formalų kontekstą

valdymo sistema. Jis skirtas paieškos kriterijams kurti, administruoti kontekstą, kurį renkami reikšmių sudarymo pagal tekstą komponentas (12 pav.). Šis įrankis leidžia redaguoti formalų kontekstą.

4.2.2. Formalaus konteksto analizė

Duomenims, kurie yra surinkti ir įrašyti į formalų kontekstą, analizuoti yra taikomas

nemokamas Toscana programinis įrankis. Šis įrankis naudoja formalių sąvokų analizę formaliam kontekstui vaizduoti braizant hierarchinį konceptų medį (5 pav.). Šio įrankio privalumas yra tas, kad jis sugeba imti duomenis iš duomenų bazių. Analizuojant programos sudaromą tekstą yra galimybė sąvokas greitai įtraukti į formalų kontekstą (6 pav.). Laukas, kuris turi nuorodą „INSERT“, reiškia, kad formaliame

OBJEKTAI	Clas	Paieskai	VIETA	KAMBARIAI	BUKLE	PASTATO_TIPAS	AUSKTAS	AUKSTU	PLOTAS	PASTATYTA	Marijampole	Vilnius	Kaunas	1_kambarys	2_kambariai
_1_kambarys	1 k b	X		X										X	
_2_kambariai	1 k	X		X										X	
_3_kambariai	1 k b.	X		X										X	
_4_kambariai	1 kamb	X		X										X	
_5_kambariai	1 kamb	X		X										X	
_bendrabutis	1 KAMBARIO	X		X										X	
_Daline apdaila	1,5 kamb.	X		X											X
_daugiau_5	1-o k.	X		X										X	
_garazas	1-o kambario	X		X										X	
_isnuomoti	1,5 kambario	X		X											X
_Kaunas	1k buta	X		X										X	
_keisti	1k,	X		X										X	
_Marijampole	1_kambarys	X		X										X	
_namas	2 k b	X		X											X
_naujas	2 k,	X		X											X
_Parduoti	2 k,	X		X											X
_patalpos	2 k b.	X		X											X
_Pirkti	2 kam,	X		X											X
_reikia remonto	2 KAMARIU	X		X											X
	2 kamb	X		X											X

13 pav. Formalus kontekstas

kontekste nėra reikiamos sąvokos, ir jeigu analizuojamame tekste yra su tuo lauku susijusi sąvoka, tai ją galima įrašyti į formalųjį kontekstą. Paspaudus nuorodą „INSERT“, atsirado sąvokos kūrimo langas (13 pav.). Įtraukus sąvoką į formalų kontekstą ir paleidus teksto apdorojimo procesą, trūkstama sąvoka duomenų bazėje yra užpildoma atitinkamos reikšmės įrašais.

5. Išvados ir rezultatai

Atliekant eksperimentą pirmu būdu, terminai (objektai) buvo surinkti ir patikslinti naudojant surinktus duomenis per vieną savaitę.

Atliekant eksperimentą taikant siūlomą metodą, terminai (objektai) buvo surinkti

ir patikslinti, surinkti duomenys įrašyti per 3 valandas.

Pastaba: į įvertintą laiką nebuvo įtrauktas programinio įrankio „Nekilnojamojo turto paieškos sistemos“ kūrimo laikas, nes šis įrankis skirtas formalaus konteksto redagavimui.

Atlikus eksperimentą nustatyta, kad siūlomas metodas leidžia greitai surasti terminus (objektus), analizuoti daug informacijos. Naudojant formaliame kontekste sukauptus duomenis galima vaizdžiai parodyti konceptų hierarchinę struktūrą bei vėliau ją remiantis sudaryti ontologiją. Tai padeda surasti priklausomybes, atitikimus, pasikartojančias struktūras, išimtis ir t. t.

LITERATŪRA

GANTER, B.; WILLE, R. (1999). Formal Concept Analysis: Mathematical Foundations. Springer, Berlin-Heidelberg.

CUNNINGHAM, H. (1999). Information Extraction – a User Guide (Second Edition)// Institute for Language, Speech and Hearing (ILASH), and Department of Computer Science. University of Sheffield, UK, <<http://www.dcs.shef.ac.uk/~hamish>> [žiūrėta 2007 m. spalio 21 d.].

DAU, F. (2004). Types and Tokens for Logic with Diagrams In K. E. Wolff, H. Pfeiffer, & H. Delugach (Eds.), Conceptual Structures at Work: 12th International Conference on Conceptual Structures. Berlin: Springer, pp. 62–93.

GRUBER, T. R. (1993). A translation approach to portable ontology specifications. Knowledge Acquisition, vol. 5, No. 2, pp. 199–220.

HUFFMAN, S. (1996). Learning information ex-

traction patterns from examples// In Wermter, Riloff, and Scheler, editors, Connectionist, Statistical, And Symbol Approaches to Learning for Natural Language Processing, volume 1040 of Lecture Notes in Artificial Intelligence, pp. 246–260, Berlin, Springer.

MARTIN, J.; ODELL, J. (1994). Object-Oriented Methods: A Foundation. Prentice-Hall, pp. 52.

WILLE, R. (1982). Restructuring Lattice Theory: an approach based on hierarchies of concept. Ordered sets / editor I. Rival. – Reidel, Dordrecht-Boston, pp. 445–470.

WORMUTH B.; BECKER P. (2004). Introduction to Formal Concept Analysis, 2nd International Conference of Formal Concept Analysis February 23 – February 27, Sydney, Australia. Prieiga per internetą: <http://www.wormuth.info/ICFCA04/Introduction_to_FCA_ICFCA2004.pdf> [žiūrėta 2007 m. rugsėjo 14 d.].

ONTOLOGY BUILDING USING FORMAL CONCEPT

Darius Jurkevičius, Olegas Vasilecas, Algirdas Laukaitis

S u m m a r y

In this article the method how to make ontology using formal concepts is described. This method is used to collect terms of specific domain. This method allows to represent the hierarchical tree of concepts without analysing terms of specific domain. The main idea of proposed method is to select terms (objects) from collected data using templates. Later collected terms are analysed us-

ing formal concept analysis method. This allows to simplify the steps of terms analysis and ontology representing in ontology creating process. In this article the experiment and it's results and conclusions are described. The realty domain is select for the experiment.

Keywords. Ontology, formal concept analysis, formal concept, programme agent.

Įteikta 2008 m.