

KAUNO TECHNOLOGIJOS UNIVERSITETAS  
VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

ALGIRDAS TRUKYS

**VEIKLOS FIDYNO IR VEIKLOS TAISYKLI  
SEMANTIKA GRINDIAMOS ONTOLOGIJ UFIKLAUSOS**

Daktaro disertacijos santrauka  
Technologijos mokslai, informatikos inžinerija (07T)

2017, Kaunas

Disertacija rengta 2016-2017 metais Kauno technologijos universiteto Informatikos fakultete, Informacijos sistemų katedroje. Mokslinius tyrimus rėmė Lietuvos mokslo taryba.

**Mokslin vadov :**

Prof. dr. Lina NEMURAITI (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija, OTT).

**Redagavo:** Antony Richard Bexon, Vilija Čeliešienė .

**Informatikos inžinerijos mokslo krypties disertacijos gynimo taryba:**

prof. dr. Robertas DAMAŠKUS (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija OTT) – **pirmininkas**;

prof. dr. Romas BARONAS (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija OTT);

doc. dr. Nikolaj GORANIN (Vilniaus Gedimino technikos universitetas, technologijos mokslai, informatikos inžinerija OTT);

prof. dr. Audrius LOPATA (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija OTT);

doc. dr. Raimundas MATULEVIČIUS (Tartu universitetas, Estija, technologijos mokslai, informatikos inžinerija OTT).

Disertacija bus ginama viešame informatikos inžinerijos mokslo krypties disertacijos gynimo tarybos posėdyje 2017 m. birželio 23 d. 11 val. Kauno technologijos universiteto disertacijų gynimo salėje.

Adresas: K. Donelaičio g. 736403, 44249 Kaunas, Lietuva.

Tel. (370) 37 300 042; faks. (370) 37 324 144; el. pašto doktorantura@ktu.lt

Disertacijos santrauka išleista 2017 m. gegužės 23 d.

Su disertacija galima susipažinti internetinėje svetainėje <http://ktu.edu> ir Kauno technologijos universiteto bibliotekoje (K. Donelaičio g. 20, 44239 Kaunas).

## TERMINAI IR SANTRUMPOS

<b>Terminas</b>	<b>Aprašymas</b>
ATL	<i>Eclipse</i> aplinkoje veikianti transformavimo kalba, vartojama modeliais grindfliamoje inflinerijoje.
BPMN	Standartas, skirtas vaizduoti veiklos proces modelius grafiškai.
EBNF	Sintaks s formalaus aprašymo notacija.
Efektyvumas	Nat ralisios kalbos s saj tinkamumo matavimo kriterijus, i-rei-kiamas tikslumo ir i-samumo matais (angl. <i>effectiveness</i> ).
Leksikonas	Nat ralisios kalbos s saj os fiodynas, naudojamas formuluoti, analizuoti ir transformuoti klausimus.
NKS	Nat ralisios kalbos s saja.
OMG	Konsorciumas, kuriantis ir palaikantis vair spektr technologini standart .
OWL	W3C konsorciumo patvirtinta tinklo ontologij kalb -eima.
Tinkamumas	Terminas, kuris apibr flia bet kokio artefakto tinkamum naudoti j pagal paskirt (angl. <i>usability</i> ).
Perkeliamumas	Savyb , leidflianti pritaikyti nat ralisios kalbos s saj skirtingoms dalykin ms sritims (angl. <i>portability</i> ).
Priimtinumas	Terminas, kuris apibr flia, kaip lengvai ir nat raliai naudotojas gali pateikti klausimus naudodamas nat ralisios kalbos s saj .
RDF	Duomen modelis, skirtas apra-yti ontologijos resursus.
SBVR	OMG konsorciumo sukurtas standartas, skirtas apra-yti veiklos fiodynus ir veiklos taisykles strukt rizuota nat rali ja kalba.
SPARQL	Semantinio tinklo ir ontologij ufklaus kalba.
W3C	Pagrindin tarptautin saityno standart k rimo organizacija.

XMI	OMG konsorciumo sukurtas metaduomen apsikeitimo standartas. Dafniausiai naudojamas programin s rangos komponent modeliams perduoti.
Xtext	Programavimo ir dalykin s srities kalb k rimo karkasas.

## 1. VADAS

Tradicinis raktiniais žodžiais grindžiamas interneto paieškos sistemos negali suprasti internete publikuojamų duomenų semantikos ir visada grąžinti tikslius rezultatus. Ufklaus rezultatai dažnai būna pertekliniai ir naudotojams tenka patiemis ieškoti reikalingos informacijos tarp pateiktų rezultatų.

Semantinio tinklo idėja remiasi tuo, kad kompiuterinės sistemos supranta internete saugomos informacijos prasmę. Semantinio tinklo pagrindas yra ontologijos, kuriose saugomos esybės, atitinkančios realaus pasaulio objektus (pvz., asmenys, transporto priemonės, organizacijos ir kt.), jų ryšiai (pvz., organizacijos, kuriose dirba asmenys ir kt.), savybės. Paieška ontologijose vadinama semantine. Ufklaus raktiniais žodžiais grindžiamai pranašesnė tuo, kad geba suprasti ufklaus prasmę ir grąžinti tikslesnius rezultatus.

Vienas iš didžiausių sunkumų kuriant semantines paieškos sistemas ó patogiai ir praktiškai naudotojų sąjau. Buvo pasiūlyta vairių naudotojų sąjų ontologijoms ó nuo paprasčiausių, skirtų ufražyti ir vykdyti *SPARQL* ufklaušas, iki leidžiančių pateikti klausimus natūralia kalba. *E. Kaufmanno* ir *A. Bernsteino* atliktas naudotojų sąjų ontologijoms tyrimas [11] parodė, kad patogiausios naudotojams yra natūraliosios kalbos sąjau (NKS).

### 1.1. Motyvacija

Yra darbe nuspręsta kurti naują semantinę paieškos sprendimą, leidžiantį pateikti klausimus natūralia kalba. Pirmoji priežastis buvo poreikis ražyti klausimus vairiomis kalbomis. Daugiakalbiškumas yra svarbus, nes 25,9 % interneto naudotojų vartoja anglų kalbą, o likę naudotojai ó kitas kalbas [15]. Nors esamos natūraliosios kalbos sąjau gerai atsako klausimus anglų kalba, jų autoriai neanalizuoja galimybių pritaikyti šias sistemas kitoms kalboms. Nėra atskleidžiama, kurie komponentai priklausomi nuo kalbos ir turėtų būti pakeisti ir kokių programinio kodo modifikacijų reikia.

Antroji priežastis yra susijusi su klausimų ir ontologijos susiejimu. Tai yra viena iš pagrindinių NKS funkcijų, kuri reikia atlikti transformuojant klausimus ufklaušas. Ontologijos projektuojamos taip, kad jose saugojama informacija būtų suprantama kompiuteriui, todėl ji struktūrai gali skirtis nuo to, kaip žmonės formuluoja klausimus (pvz., dauginamieji ryšiai atvejais). Dėl šios priežasties NKS turi sugebėti susieti klausimus ne tik su vienu, bet ir su grupe ontologijos resursų. Esant NKS analizė atskleidė, kad šios sistemos leksikonai dažniausiai sudaro automatiškai pagal ontologijos struktūrą ir leidžia pateikti tik tokius klausimus, kurių formuluočiai atitinka ontologiją.

Buvo suformuluoti šie pagrindiniai kuriamos sistemos architektūros principai: (1) sistema turi būti pritaikoma pateikti klausimus skirtingomis kalbomis; (2) leksikonas privalo leisti susieti paprastus klausimus su sudėtingomis ontologijos struktūromis. Sistemos pagrindu nuspręsta naudoti *SBVR* standartą, kuris skirtas apražyti veiklos žodynų ir veiklos taisyklės struktūrizuota kalba bei leidžia pateikti

klausimus programin s rangos modeliams. Kadangi *SBVR* metamodelis atskiria prasm nuo vaizdavimo, veiklos flodyno konceptai, taisykl s ir klausimai gali b ti i-reik-ti skirtingai ó taip pat ir skirtingomis kalbomis. Tai leidffia daryti prielaid , kad klausimai gal t b ti transformuojami uflklausas naudojant nuo kalbos nepriklausomas transformavimo taisykles.

*SBVR* metamodelis leidffia apra-yti flodyno koncept apibr fimus ir i-vesti koncept prasm naudojant kitus konceptus. Tod l *SBVR* i-vedimo taisykl s gal t b ti naudojami susieti paprastus naudotojo pateiktus klausimus su ontologijos resurs kombinacijomis.

Nors *SBVR* standartas leidffia formuluoti klausimus programin s rangos modeliams, iki -iol jis nebuvo naudojamas semantinei paie-kai. Tš darbas tur t atsakyti, ar *SBVR* standartas gali b ti naudojamas semantin s paie-kos sprendimui ir ar jis leidffia pasiekti daugiakalbi-kum bei susieti paprastus klausimus su sud -tingomis ontologijos strukt romis.

## 1.2. Tyrimo sritis ir objektas

Tš tyrimo objektas yra ontologij uflklaus vykdyimo procesas, pateikiant klausimus nat rali ja kalba. Tyrimo sritis apima -ias temas:

- nat ralisios kalbos s sajos, skirtos flini baz ms;
- pafangiausi flini ir duomen modeliai, j vaizdavimo ir uflklaus kalbos (*SBVR*, *OWL 2*, *RDF*, *SPARQL*), susij rankiai ir technologijos;
- modeliais grindffiamos transformacij technologijos.

## 1.3. Sprendffiama problema ir keliami klausimai

Tš tyrimas paskatintas patogi ir prakti-k ontologij s sajam, leidffian i pateikti klausimus nat rali ja kalba, tr kumo. Sprendffiant min t problem buvo siekiama vykdyti -iuos reikalavimus:

- galimyb pateikti klausimus vairiomis kalbomis (-iame darbe nagrin jamomis angl ir lietuvi bei gramati-kai pana-iomis: vokie i , ek , lenk ir kt.);
- sprendimas turi leisti susieti klausimus su ontologijos resurs kombinacijomis;
- perkeliamumas;
- atsakym klausimus efektyvumas turi b ti pana-us kit NKS.

Tyrimas turi atsakyti -iuos klausimus:

1. Ar *SBVR* klausimai gali b ti naudojami vykdyti ontologijos uflklausas ir susieti klausimus su ontologijos resurs kombinacijomis?
2. Kaip galima transformuoti nat ralisios kalbos klausimus *SPARQL* uflklausas naudojant *SBVR*?
3. Ar manoma sukurti perkeliam ir kartu teisingai klausimus atsakan i nat ralisios kalbos s sajam ontologijoms, naudojant *SBVR*?
4. Ar galima *SBVR* grindffiam nat ralisios kalbos s sajam ontologijoms pritaikyti skirtingoms kalboms ir kurios sistemos dalys priklauso nuo kalbos?

#### 1.4. Tikslai ir uždaviniai

Tyrimo tikslas – išplėsti semantinių paieškų galimybes, leidžiant naudotojams pateikti natūraliosios kalbos klausimus skirtingomis kalbomis bei tokius klausimus, kurių susiejimas su ontologija sudėtingas (t. y., klausimai turi būti susieti su ontologijos resursų kombinacijomis). Tyrimo uždaviniai yra šie:

1. išanalizuoti literatūrą, susijusią su OWL 2 ontologijomis ir ontologijų ufklausų kalba SPARQL; esamas NKS, skirtas duomenų bazėms ir ontologijoms; SBVR finišo modelis ir SBVR specifikacijų kūrėjų rankius;
2. apibrėžti kuriamos NKS koncepciją ir algoritmus, leidžiančius transformuoti natūraliosios kalbos klausimus SPARQL;
3. apibrėžti SBVR rankio, skirtą specifikuoti veiklos flūdynus, taisykles ir pateikti klausimus, koncepciją;
4. sukurti prototipus sprendimo tinkamumui vertinti;
5. atlikti eksperimentus ir vertinti tyrimo rezultatus.

Šis darbas gali išspręsti tik dalį problemų, susijusių su semantinių paieškų poreikiu. Pagrindiniai kokybės kriterijai šie:

- galimybė pateikti klausimus skirtingomis kalbomis;
- galimybė susieti klausimus su ontologijos resursų kombinacijomis;
- sprendimo perkeliamumas;
- atsakymų klausimus efektyvumas.

#### 1.5. Tyrimo metodika

Tyrimas buvo vykdomas pagal konstruktyviojo mokslinio tyrimo metodiką, kuria siekiama išplėsti flmogaus ar organizacijos galimybių ribas ir sukurti naują bei patį ang arto faktą [10]. Tyrime naudoti metodikos etapai pateikti disertacijoje.

#### 1.6. Ginamieji teiginiai

Darbe ginami šie teiginiai:

1. SBVR leidžia naudoti nuo kalbos nepriklausomas klausimų transformavimo semantines ufklausas taisykles ir pasiekti NKS ontologijoms daugiakalbiškumą. Tai iau sprendimui vykdyti reikia ir nuo kalbos priklausomų komponentų, skirtų atlikti sintaksinį ir morfologinį klausimų analizę.
2. SBVR išvedimo taisyklės leidžia susieti natūraliosios kalbos klausimus ir sudėtingą ontologijos struktūrą. Todėl ufklausos ontologijose gali būti vykdomos pateikiant klausimus, kurie tiesiogiai neatitinka jos struktūros.
3. SBVR grindžiama NKS yra perkeliama. Perkeliamumas uftikrinamas aprašant dalykinės srities flūdynes ir taisykles (t. y. leksikoną) ir susiejant jį su ontologija. Susiejimas atliekamas paflymint ontologijos resursus atitinkanti flūdyno koncepcijai.

#### 1.7. Asmeninis indėlis ir naujumas

Pagrindinis šio darbo indėlis ó semantinių paieškų sprendimas, leidžiantis

pateikti klausimus nat rali ja kalba. Sprendimas sudarytas i–dviej dali :

- Ontologijoms skirta NKS, kuri geba pateikti ir analizuoti nat raliuos kalbos klausimus bei transformuoti juos *SPARQL* ufklausas.
- *SBVR* strukt rizuotos kalbos ra–ykl , naudojama konfig ruoti NKS. Ra–ykl leidffia specifikuoti veiklos flodynus, veiklos taisykles ir pateikti klausimus strukt rizuota kalba. Specifikacijos gali b ti transformuojamos *SBVR* modelius *XMI* formatu.

Tyrimo naujumas:

1. Nebuvo rasta kit autori tyrim , kuriuose *SBVR* klausimai b t naudojami ontologijos ufklausoms vykdyti.
2. Sukurta NKS turi ai–kiaiai apibr fttas dalis, kurias reikia pakeisti, norint pritaikyti sistem klausin ti skirtingomis kalbomis.
3. Sukurta NKS galima pateikti klausimus, kai ontologijos strukt ra yra sud tinga ir klausimus reikia susieti su ontologijos resurs kombinacijomis. Susiejimas apra–omas naudojant *SBVR* i–vedimo taisykles arba formalius *SBVR* koncept apibr flimus.
4. Sukurta *SBVR* strukt rizuotos kalbos ra–ykl , skirta ra–yti veiklos flodyn ir veiklos taisykli specififikacijos, naudojama formuluoti klausimus. rankis leidffia padalinti specififikacijos atskiras dalis ir pasl pti tik i–vedimui naudojamus flodyno konceptus nuo naudotojo. TŲra–ykl s savyb taip pat leidffia kurti metaflodynus ir pritaikyti rank kitiems tikslams (pvz., transformacijoms *OWL 2* ontologijas, *BPMN* veiklos proces specifikavimui ir kt.).

## 1.8. Praktin reik–m

Sprendim galima pritaikyti vykdant paie–k skirting dalykini sri i ontologijose ir pateikiant klausimus skirtingomis kalbomis. *SBVR* i–vedimo taisykl s leidffia pateikti paprastus klausimus sud tingoms ontologijoms, kurios dafnai naudojamos praktikoje. Sprendim galima naudoti semantinei paie–kai internete bei verslo taikomosiose programose.

TŲb darbo rezultatai pritaikyti vykdant projekt ųSemantika LTų [18], suk rus semantin s paie–kos paslaug , kuri leidffia pateikti klausimus lietuvi kalba. *SBVR* strukt rizuotos kalbos ra–ykl buvo naudojama sukurti *SBVR* veiklos flodynus ir veiklos taisykles bei pritaikyti NKS analizuotoms dalykin ms sritims (t. y., politikos, ekonomikos ir vie–ojo administravimo). Darbe sukurtos transformacijos buvo naudotos transformuoti lietuvi kalbos klausimus *SPARQL* ufklausas, skirtas vykdyti paie–k semanti–kai anotuotuose lietuvi–k naujien portal tekstuose.

Sukurta *SBVR* strukt rizuotos kalbos ra–ykl gali b ti naudojama kurti *SBVR* specififikacijos kituose, su *SBVR* susijusiuose tyrimuose.

## 1.9. Rezultat aprobavimas

Tyrim rezultatai buvo pristatyti penkiose tarptautin se konferencijose ir vie–noje Lietuvos konferencijoje. Du straipsniai publikuoti mokslo flurnaluose, kurie



referuojami Mokslin s informacijos instituto *ISI* pagrindiniame s ra-e su citavimo indeksu. Du straipsniai paskelbti referuojamuose *ISI* konferencij leidiniuose. Keturi straipsniai paskelbti kituose mokslo leidiniuose. I-samus publikacij s ra-as pateiktas skyriuje šAutoriaus publikacij disertacijos tema s ra-aso.

### 1.10. Disertacijos strukt ra

Antrajame disertacijos skyriuje pristatomos ontologijos, NKS analiz , kurioje atskleidfiami pagrindiniai toki kuriam sistem reikalavimai ir sunkumai, su kuriais susiduriama. Tame skyriuje taip pat pateikiama *SBVR* specifikacijos ir *SPARQL* uflklaus kalbos analiz .

Tre iajame skyriuje pristatomas semantin s paie-kos sprendimas, pateikiami darbe sukurti modeliai ir algoritmai.

Ketvirtajame skyriuje apra-oma sukurt prototip realizacija.

Penktajame skyriuje pateikiami eksperimentai, skirti vertinti sukurt sprendim .

Darbo apibendrinimas ir i-vados pateikiamos -e-tajame disertacijos skyriuje. Po i-vad eina darbe naudotos literat ros ir autoriaus publikacij darbo tema s -ra-ai.

## 2. NAT RALIOS KALBOS S SAJ IR PAFiANGIAUSI fiINI MODELI ANALIZ

### 2.1. Ontologijos apibr fimas

Kompiuteri ir informacijos moksluose ontologija yra techninis terminas, rei-kiantis artefakt , skirt modeliuoti tam tikros dalykin s srities finias [9]. 1993 metais *T. Gruberis* termin *ontologija* apibr fl kaip *tiksl i konceptualizavimo specifikacij* [8].

Kompiuteriui suprantamoms ontologijoms kurti naudojama tinklo ontologij kalba *OWL*, kuri leidfia apra-yti klases, savybes bei egzempliorius [19]. *OWL* ontologijos grindfiamos *RDF* duomen modeliu, t. y., grafu, sudarytu i- trilypi (angl. *triple*) formuluo i , i-reik-t subjekto, predikato ir objekto pavidalu.

Tame darbe semantinei paie-kai vykdyti naudojamos dalykin s srities (angl. *domain*) ontologijos, specifikuotos *OWL* kalba (naujausia jos versija *OWL 2*). Pagal semantin i-rai-kingum gali b ti naudojama *OWL Lite* arba *OWL DL* kalbos versija.

### 2.2. Nat raliuosios kalbos s sajos

Analiz atskleid svarbiausias savybes, kurias siekiama uftikrinti kuriant NKS ontologijoms: perkeliamumas (angl. *portability*), kuris leidfia pritaikyti sistem vairioms dalykin ms sritims ir priimtinumas (angl. *habitability*), kuris apibr fia, kaip lengvai ir nat raliai naudotojas gali pateikti klausimus [22]. Priimtimumas tiesiogiai susij s su bendresne s voka ó tinkamumu (angl. *usability*) [5], kurio vienas i-matavimo kriterij yra efektyvumas (angl. *effectiveness*) [2].

I-analizavus esamas NKS ontologijoms (*QuestIO* [6], *FREyA*, [4] *ORAKEL*

[3] *PANTO* [21], *Querix* [12]), paaiškio, kad šios sistemos yra perkeliamos, o priimtinumas padidinamas naudojant vairius metodus o kontekstin menu, patikslinimo dialog ir kt. Pagrindiniai ši sistem tr kumai, kuriuos siekiama išspr sti darbe, yra šie: n ra galimyb s pateikti klausimus skirtingomis kalbomis bei tokius klausimus, kuri strukt ra tiesiogiai neatitinka ontologijos.

### **2.3. SBVR standartas ir jo taikymas semantini s paieškos sprendimui**

*SBVR* yra *OMG* konsorciumo specifikacija [17], skirta kurti veiklos flodyn ir veiklos taisykli specifikacijas bei aprašyti dalykin s srities semantik strukt rizuota nat rali ja kalba.

Pirmoji prieflastis, d l kurios nuspr sta *SBVR* naudoti semantini s paieškos sprendimui, yra šio standarto savyb atskirti prasm (angl. *meaning*) nuo vaizdavimo (angl. *representation*). Tš dali atskyrimas leidžia prasm išreikšti vairiais b dais (pvz., tekstu, simboliais, garsais ir kt.) bei sudaro galimyb pasiekti daugialbi-kum [16].

Kita prieflastis o išvedimo taisykl s (angl. *derivation rules*), kurios leist susieti nat rališios kalbos klausim formulutes su ontologijos strukt ra.

### **2.4. Ontologij ufklaus kalba SPARQL**

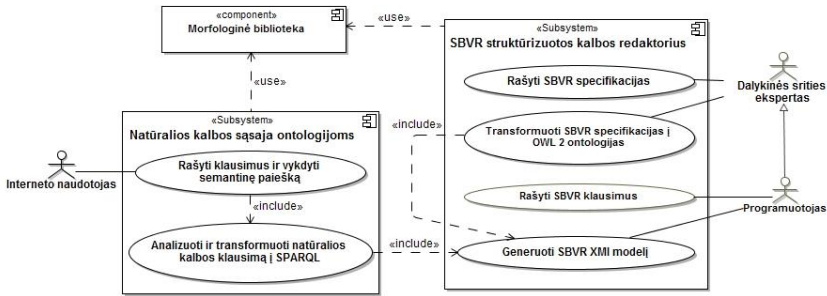
*SPARQL* o *W3C* standartas ir *de facto* ufklaus kalba ontologijoms [20]. *SPARQL* grindžiama trilypi formuluo išablono (angl. *triple pattern*) ir *RDF* grafo atitikimo tikrinimu.

Buvo vykdoma *SPARQL* sintaks s metamodelio analiz , kuri atskleid , kad *SPARQL* sintaks s ir *SBVR* klausim metamodeliai turi atitinkan iš element (pvz.: *SBVR* atomini formuluo iš veiksmaflo diniai konceptai atitinka *SPARQL* trilypi formuluo iš ablonus), tod l *SBVR* klausim modeliais grindžiama transformacija *SPARQL* ufklausas yra manoma.

## **3. VEIKLOS ŠODYNO IR VEIKLOS TAISYKLI SEMANTIKA GRINDŽIAMAS SEMANTINI S PAIEŠKOS SPRENDIMAS**

### **3.1. Semantini s paieškos sprendimo reikalavimai**

Semantini s paieškos sprendim sudaro NKS ontologijoms ir *SBVR* strukt rizuotos nat rališios kalbos rašykl . Funkciniai sprendimo reikalavimai pateikti 3.1 pav. NKS ontologijoms skirta interneto naudotojams, kurie gali pateikti klausimus nat rali ja kalba bei vykdyti semantini paiešk *OWL 2* ontologijose. *SBVR* rašykl s naudotojai yra dalykin s srities ekspertas ir programuotojas. Dalykin s srities ekspertas kuria veiklos flodyn ir veiklos taisykli specifikacijas bei konfig ruoja NKS. Programuotojas rašykl naudoja rašyti klausimus strukt rizuota nat rali ja kalba ir generuoti *SBVR XMI* modelius. Tš modeliai naudojami kurti ir testuoti klausim *SPARQL* ufklausas transformacijas.



3.1 pav. Semantin s paie-kos sprendimo funkciniai reikalavimai

SBVR struktūrizuotos kalbos rašyklė ir NKS ontologijoms naudoja internetinę morfologinę biblioteką, skirtą analizuoti žodžius (t. y., rasti lemas ir kitas morfologines savybes) bei generuoti reikiamą morfologinių formų žodžius.

### 3.2. SBVR struktūrizuotos kalbos rašyklė

SBVR struktūrizuotos kalbos rašyklės koncepciją sudaro gramatikos taisyklės, skirtos specifiškai veiklos žodynui, veiklos taisyklės ir pateikti klausimai. Gramatiką sudaro leksikos ir generavimo taisyklės. Pagrindinės leksikos taisyklės pateikiamos 3.1 lentelėje.

3.1 lentelė. SBVR struktūrizuotos kalbos rašyklės leksikos taisyklės

```
terminal TERM_OR_VERBSYMBOL:
  LOWERCASE (LOWERCASE | DIGIT | '_' | '-' ) * ;
terminal NAME:
  (UPPERCASE | DIGIT) (LOWERCASE | UPPERCASE | DIGIT | '_' | '-' ) * ;
terminal NEWLINE:
  ('\r' | '\n') * ;
terminal ML_COMMENT:
  '/' * -> '* /' ;
terminal fragment LOWERCASE:
  ('a' .. 'z') | 'a' | 'č' | 'ę' | 'ė' | 'į' | 'š' | 'u' | 'ū' | 'ž' ;
terminal fragment UPPERCASE:
  ('A' .. 'Z') | 'Ą' | 'Ć' | 'E' | 'Ė' | 'Į' | 'Š' | 'Ū' | 'Ū' | 'Ž' ;
terminal fragment DIGIT:
  ('0' .. '9') ;
```

Smulkausios generavimo taisyklės, skirtos apibrėžti pagrindinius SBVR sintaksinius elementus, pateiktos 3.2 lentelėje.

3.2 lentelė. Generavimo taisyklės, skirtos apibrėžti pagrindinius SBVR sintaksinius elementus

```
Term:
  syntax=TERM_OR_VERBSYMBOL;
Name:
```

```

syntax=NAME;
VerbSymbol:
syntax=TERM_OR_VERBSYMBOL;

```

Veiklos fodyno gramatikos pagrind sudaro *SBVR* koncept specifikuojimo taisyklės, kurios pateiktos 3.3 lentelėje.

**3.3 lentelė**. Pagrindinės veiklos fodyno gramatikos taisyklės

```

GeneralConcept:
  primaryRepresentation=Term
  (captions+=CaptionForGeneralConcept)*;
IndividualConcept:
  primaryRepresentation=Name
  (captions+=CaptionForIndividualConcept)*;
VerbConcept:
  primaryRepresentation=VerbConceptWording
  (captions+=CaptionForVerbConcept)*;

```

3.4 lentelėje pateiktos pagrindinės veiklos taisyklės gramatikos taisyklės. Gramatika leidžia specifikuoti struktūras, operacines ir išvedimo taisykles. Veiklos taisyklės aprašomos naudojant atomines formules ir reikinius, kurie taip pat pateikti 3.6 lentelėje.

**3.4 lentelė**. Pagrindinės veiklos taisyklės gramatikos taisyklės

```

Rule:RuleBasedOnStatementOfAtomicFormulation|ImplicationRule;
RuleBasedOnStatementOfAtomicFormulation:
  modality=ModalOperator
  statement=RepresentationOfAtomicFormulation;
ImplicationRule:
  modality=ModalOperator
  consequent=RepresentationOfAtomicFormulation
  antecedent=RepresentationOfAtomicFormulation;

```

Pagrindinės klausimų gramatikos taisyklės pateiktos 3.5 lentelėje. Taisyklės leidžia atpažinti dvejį tipų klausimus ir skirtus surasti nurodyto tipo objektus ir sudarytus remiantis atomine formule ir reikiniiais.

**3.5 lentelė**. Klausimų taisyklės

```

Question:
  (QuestionToFindInstancesByType|OrdinaryQuestion);
QuestionToFindInstancesByType:
  startKeyword= FIND
  concept=[Term];
OrdinaryQuestion:
  startKeyword=(WH|FIND)

```

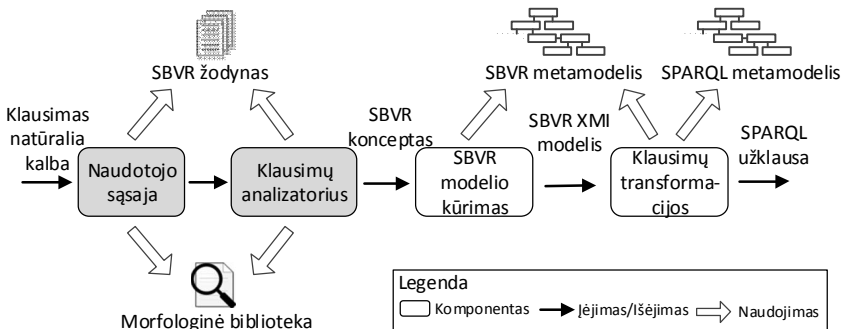
```
statement=RepresentationOfAtomicFormulation
("?"|".");
```

### 3.6 lentelė. Pagrindinis atominių formuluočių reikšmės taisyklės

```
RepresentationOfAtomicFormulation:
BothPlaceholdersNotReplaced|
1stPlaceholderReplacedByName|
2ndPlaceholderReplacedByName|
BothPlaceholdersReplacedByName|
2ndPlaceholderReplacedByQuantityRestriction|
2ndPlaceholderReplacedByQuantification;
```

### 3.3. Natūraliosios kalbos sąjona ontologijoms

NKS ontologijoms koncepcija pateikta 3.2 pav. Tamsia spalva pažymėti nuo kalbos priklausomi komponentai (t. y., pasiūlymų sudarymas vedant klausimus naudotojo sąjonoje ir euristinis klausimų analizės taisyklės). Taip pat nuo kalbos priklausoma morfologinė biblioteka.



3.2 pav. NKS komponentai

Naudotojo sąjona yra skirta rašyti klausimus ir pateikti rezultatus. Šis komponentas naudoja žodžių pasiūlymų ir automatinio uždavinio (angl. *autocomplete*) mechanizmą ir padeda rašyti klausimus pažodžiui. Žodžių pasiūlymai sudaromi remiantis SBVR veiklos žodynų daiktavardiniais ir veiksmažodžiniais konceptais. Pagal SBVR veiklos žodyne aprašytą konceptų hierarchiją, gali būti formuluojami specifiniai ar abstraktūs klausimai, SBVR sinonimai leidžia rašyti vairesnius klausimus. Jeigu sistema negali vienareikšmiškai suprasti klausimo, naudotojui rodomas patikslinantis dialogas, kuriame prašoma pateikti teisingą interpretaciją.

Kai ontologijos struktūra sudėtinga ir tiesiogiai neatitinka klausimų, konceptai, naudojami formuluojant žodžius klausimus, aprašomi formaliais SBVR apibrėžimais arba išvedimo taisyklėmis.

Naudotojo vestas klausimas analizuojamas vykdant sintaksinį ir morfologinį analizę. Analizės tikslas – nustatyti klausimų sudarant SBVR daiktavardinę ar

veiksmafodin koncept , pagal kur *SBVR* modelio k rimo komponentas suformuoja klausimo *SBVR* model *XMI* formatu. Tš modelis naudojamas vykdant modeliais grindfiamas transformacijas *SPARQL*. Transformacij taisykl s pateiktos disertacijoje.

## 4. SEMANTIN S PAIETKOS SPRENDIMO REALIZACIJA

### 4.1. *SBVR* strukt rizuotos kalbos ra-ykl s realizacija

*SBVR* strukt rizuotos kalbos ra-ykl s prototipas realizuotas naudojant *Xtext* karkas [7]. Ra-ykl atpafli sta tekstus, para-ytus pagal gramatikoje apibr ft sintaks bei turi -ias funkcijas: *SBVR* koncept sintaks s ir susiejimo tikrinimas; gramatikos element flym jimas ir spalvinimas *SBVR* stiliumi; ra-ykl s pritaikymas vairioms kalboms, integruojant morfologin bibliotek ; galimyb i-pl sti ra-ykl , papildant gramatik naujomis taisykl mis; galimyb i-skaidyti *SBVR* specifikacijas atskiras dalis; *SBVR* modelio generavimas *XMI* formatu; bibliotekos, skirtos transformuoti *SBVR* *OWL 2*, integravimas.

### 4.2. Nat ralisios kalbos s sajos ontologijoms realizacija

NKS realizacij sudaro -ios dalys:

- grafin naudotojo s saja, skirta pateikti klausimus nat rali ja kalba;
- komponentas, skirtas patikrinti klausimus ir sudaryti si lom flodfli s ra- vedant klausimus;
- komponentas, skirtas transformuoti klausimus *SPARQL* uflklausas;
- komponentas, skirtas vykdyti uflklausas ir formuoti rezultatus;
- semantinio anotavimo komponentas, skirtas ufpildyti ontologij eg-zemplioriais.

Reikia pabr fti, kad naudotojo s saj , uflklaus vykdymo ontologijoje ir rezultat formavimo bei semantinio teksto anotavimo komponentus realizavo bendraautoriai, vykdantys projekt šSemantika LTō [18]. Tš darbo rezultatai realizuoti kaip tinklo paslaugos (angl. *Web Service*), skirtos tikrinti klausimus ir sudaryti si lom flodfli s ra- bei transformuoti klausimus *SPARQL* uflklausas. Klausim transformacijos realizuotos *ATL* kalba.

## 5. EKSPERIMENTINIS TYRIMAS

*SBVR* strukt rizuotos kalbos ra-ykl s eksperimentinis tyrimas buvo vykdomas siekiant i-siai-kinti, ar: (1) sukurta gramatika ir ra-ykl tinkami kurti *SBVR* specifikacijas ir palaiko visas reikiamas *SBVR* konstrukcijas; (2) leidfia kurti *SBVR* specifikacijas skirtingose dalykin se srityse ir uftikrinti NKS perkeliamum ; (3) leidfia kurti *SBVR* specifikacijas skirtingomis kalbomis ir uftikrinti NKS daugiakalbi-kum . Ra-ykl s eksperimentinis tyrimas parod , kad:

- sukurta *SBVR* strukt rizuotos kalbos gramatika leidfia naudoti visas reikiamas veiklos flodyno konstrukcijas; specifiukuoti strukt rines, operacines ir i-vedimo taisykles, naudojant visus apibr ftus atomini formuluo i

rei-kinius; ra-yti klausimus, naudojant visus apibrėžtus atominių formuluočių rei-kinius;

- ra-yti sudaro s lygias uftikrinti konfigūruojamos NKS perkeliama ir daugiakalbi-kum, nes leidžia ra-yti SBVR specifikacijas viariose dalykinėse srityse ir skirtingomis kalbomis, pritaikius morfologinį bibliotekos ir morfologinį generatorių.

NKS eksperimentinis tyrimas buvo vykdomas siekiant vertinti sukurto sprendimo efektyvumą, perkeliama ir daugiakalbi-kum. Eksperimentui buvo naudojama Ray Mooney flinė bazė ir klausimų rinkinys [14], dafnai naudojami vertinant tokio tipo sistemas. Tai leido palyginti sukurtą sprendimą su kitais sprendimais. Taip pat atliktas eksperimentas siekiant vertinti galimybes vykdyti uftiklausas, kai klausimų susiejimas su ontologija nėra paprastas. NKS eksperimentinis tyrimas parodė, kad:

- semantinės paieškos sprendimas yra perkeliama ir j galima naudoti klausimams pateikti viariose dalykinėse srityse. Tačiau norint pasiekti gerą efektyvumą, gali tekti patobulinti euristines klausimų analizės taisykles, jeigu naudojamos klausimų formuluočių, su kuriomis anksčiau nebuvo susidurta;
- semantinės paieškos sprendimas gali būti pritaikomas klausimams pateikti skirtingomis kalbomis. Klausimų transformavimo taisyklės nepriklauso nuo kalbos, kuria klausimas para-ytas, nes transformacijose naudojami klausimų prasmės modeliai.
- semantinės paieškos sprendimas leidžia susieti klausimus su ontologijos resursų kombinacijomis. Susiejimas atliekamas naudojant konfigūravimo metodus para-ytas SBVR išvedimo taisykles arba formalius konceptų apibrėžimus.

## 6. ĮVADAS

1. Mokslini publikacijų analizė parodė, kad priimtinausia naudotojų sąjaja, skirta vykdyti paiešką ontologijose, yra natūraliosios kalbos sąjaja (NKS). Patys svarbiausi reikalavimai, keliami NKS, yra pritaikomumas skirtingoms kalboms, galimybė vykdyti uftiklausas net ir esant sudėtingoms ontologijos struktūroms, perkeliama ir priimtina. Esami NKS sprendimai leidžia pateikti klausimus tik anglų kalba ir dafniausiai tokius, kurių formuluočių tiesiogiai atitinka ontologijos struktūrą. Kita vertus, -ie sprendimai uftikrina perkeliama ir gali būti pritaikyti pateikti klausimus skirtingose dalykinėse srityse. Siekiant padidinti NKS priimtinumą, naudojami viariose srityse metodai, kuriuos galima skirstyti dvi grupes: (1) skirti supaafindinti naudotojus su leksikonu ir padėti formuluoti klausimus; (2) skirti padėti interpretuoti ir suprasti jau para-ytus klausimus.
2. SBVR analizė parodė, kad -io standarto metamodelio skiriamasis bruožas atskirti prasmę nuo vaizdavimo leidžia pasiekti daugiakalbi-kum. Tai leidžia

daryti prielaid , kad *SBVR* grindffiamoje NKS klausimai semantines ufklausas gal t b ti transformuojami naudojant nuo kalbos nepriklausomas taisykles, nes klausimo prasm s modelis yra tas pats visoms kalboms. Nuo kalbos priklausomi tokios NKS komponentai tur t b ti tik tie, kurie padeda ra- yti ir interpretuoti klausimus. Norint pasiekti daugiakalbi-kum , -ie komponentai tur t b ti pakei iami ar pritaikomi tam tikrai kalbai. NKS architekt ra buvo kuriama vadovaujantis -iomis id jomis. Kitas svarbus aspektas, susij s su pasirinkimu naudoti *SBVR* NKS sprendimui, yra i-vedimo taisykl s, kurios gali b ti naudojamos susieti paprastus klausimus su sud tinga ontologijos strukt ra.

3. *SPARQL* analiz atskleid , kad -ios kalbos ufklaus sintaks ir *SBVR* klausim metamodelis turi atitinkan i element , i-rei-kian i informacijos poreik ir ufklausos apribojimus. Tai leido daryti prielaid , kad *SBVR* klausim transformavimas *SPARQL* yra manomas. Buvo nuspr sta detaliai apra-yti atitikimus tarp *SBVR* klausim metamodelio ir *SPARQL* bei sukurti transformavimo taisykles.
4. Svarbi s lyga norint naudoti *SBVR* standart NKS ontologijoms yra redagavimo rankis, skirtas apra-yti *SBVR* veiklos flodynus ir taisykles (t. y., konfig ruoti NKS) bei sudaryti klausim *XMI* modelius transformavimui. Esam ranki analiz parod , kad nei vienas i-j neatitinka keliam reikalavim , o tobulinimas yra komplikuotas. Tod l buvo nuspr sta kurti nauj *SBVR* ra- ykl .
5. *SBVR* rankio koncepcija grindffiamu strukt rizuotos kalbos gramatika (apra- yta pana-ia *EBNF* forma), skirta specifikuoti veiklos flodynus, taisykles ir ra-yti klausimus. Gramatika apra-yta i-analizavus *SBVR* vaizdavimo (angl. *representation*) metamodel , strukt rizuotos kalbos pavyzdffius, pateiktus *SBVR* specifikacijoje bei naudotus praktikoje. Gramatika leidffia ra-yti klausimus, skirtus i-rikti nurodyto tipo objektus, klausimus su modifikatoriais, kardinalumo apribojimais, skaitiniais palyginimais ir skai iavimo funkcija.
6. Gramatika specifikuota ir *SBVR* ra-ykl realizuota naudojant *Xtext* karkas . Eksperimentai parod , kad rankis leidffia naudoti visas reikiamas *SBVR* konstrukcijas. Jis taip pat sukuria visas reikalingas s lygas perkeliama daugiakalbiui NKS sprendimui ir yra tinkamas konfig ravimui.
7. *SBVR* grindffiamos NKS ontologijoms koncepcija sudaryta i--i komponent (tik pirmieji du yra priklausantys nuo kalbos):

vartotojo s saja, leidffiant pateikti nat ralisios kalbos klausimus, atitinkan ius *SBVR* flodyn ;

klausim analizatorius, skirtas identifikuoti klausimo *SBVR* koncept ;  
*SBVR* modelio k rimo komponentas, skirtas sudaryti klausimo *SBVR* model *XMI* formatu;

transformavimo komponentas, skirtas transformuoti klausimo *SBVR* *XMI* model *SPARQL* ufklaus .



8. *SBVR* grindfliama NKS buvo realizuota ir tai parodė, kad *SBVR* klausimai gali būti naudojami vykdyti ontologijos ufklausas. Be to, eksperimentinis tyrimas atskleidė, kad sukurtas sprendimas leidžia vykdyti ufklausas ontologijose, kuri struktūra, išreikianti aktualias dalykinis srities finias, tiesiogiai neatitinka klausimų formuluočių. Tai pasiekama aprašius *SBVR* konceptų išvedimo taisykles, kurios naudojamos transformuojant klausimus.
9. Klausimų pateikimo eksperimentai dviejose dalykinis srityse rodė, kad galima pasiekti *SBVR* grindfliamos NKS perkeliamumą ir daugiakalbiškumą. Perkeliamumas pasiekiamas leidžiant aprašyti tam tikros dalykinis srities *SBVR* specifikaciją ir ją susieti su ontologija. Efektyvumo tyrimo rezultatai panašūs kitų sprendimų (t. y., geografijos dalykinis srityje f-matas 0,86, o restoran 0,75). Visgi pagrindiniai sukurto sprendimo privalumai yra daugiakalbiškumas ir galimybė pateikti klausimus, kuri struktūra tiesiogiai neatitinka ontologijos struktūros.

## LITERATŪRA

1. **Berners-Lee, T., Hendler, J., Lassila, O.** The Semantic Web, *Scientific American*, May 2001, 28637.
2. **Brook, J.** SUS ó A quick and dirty usability scale. P. Jordan, B. Thomas, B. Weerdmeester, A. McClelland (eds.) *Usability Evaluation in Industry*, 1996, 189-194.
3. **Cimiano, P., Haase, P., Heizmann, J., Mantel, M.** ORAKEL: A Portable Natural Language Interface to Knowledge Bases. Technical report, Institute AIFB, University of Karlsruhe (2007).
4. **Damljanovic, D., Agatonovic, M., Cunningham, H.** Natural Language Interfaces to Ontologies: Combining Syntactic Analysis and Ontology-based Lookup through the User Interaction. In *Proceedings of the 7th international conference on The Semantic Web: research and Applications*, Heraklion, Crete, Greece, May 30 - June 3, 2010, 106-120.
5. **Damljanovic, D., Bontcheva, K.** Towards Enhanced Usability of Natural Language Interfaces to Knowledge Bases. *Web 2.0 & Semantic Web*, 2009, 105-133.
6. **Damljanovic, D., Tablan, V., Bontcheva, K.** A Text-based Query Interface to OWL Ontologies. In *Proceedings of 6th Language Resources and Evaluation Conference (LREC)*, Marrakech, Marocco, May 28-30, 2008, 205-212.
7. **Eysholdt, M., Behrens, H.** Xtext: implement your language faster than the quick and dirty way. In *Proceedings of the ACM international conference companion on Object oriented programming systems languages and applications companion ( OOPSLA@2010)*, Reno/Tahoe, Nevada, USA, October 17-21, 307-309.
8. **Gruber, T.** A translation approach to portable ontologies. *Knowledge*

- Acquisition*, 5(2), 1993, 199-220.
9. **Gruber T.** Ontology. *Encyclopedia of Database Systems*, Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag US, 2009.
  10. **Hevner, A. R., March, S. T., Park, J., Ram, S.** Design Science in Information Systems Research. *MIS Quarterly*, 28 (1), 2004, 75-105.
  11. **Kaufmann, E., Bernstein, A.** Evaluating the Usability of Natural Language Query Languages and Interfaces to Semantic Web Knowledge Bases. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(4), 2010, 377–393.
  12. **Kaufmann, E., Bernstein, A., Zumstein, R.** Querix: A Natural Language Interface to Query Ontologies Based on Clarification Dialogs. In *Proceedings of 5th International The Semantic Web Conference (ISWC 2006)*, Athens, Georgia, USA, November 5-9, 2006, 9806981.
  13. **Linehan, M. H.** Semantics in Model-Driven Business Design. In *Proceedings of 2nd International Semantic Web Policy Workshop (SWPW'06)*, Athens, Georgia, November 5-9, 2006, 86-93.
  14. **Natural Language Learning Data.** Available from: <http://www.cs.utexas.edu/users/ml/nldata.html> [Accessed: 05 Apr 2015].
  15. **Number of Internet Users by Language.** Internet World Stats, Miniwatts Marketing Group, 30 November 2015. Available from: <http://www.internetworldstats.com/stats7.htm>. [Accessed: 15 Mar 2016].
  16. **Ross, R. G.** The RuleSpeak® Business Rule Notation. Available from: <http://www.brcommunity.com/b282.php> [Accessed 18 June 2013].
  17. **Semantics of Business Vocabulary and Business Rules (SBVR), Version 1.3.** OMG Document Number: formal/2015-05-07, 2015.
  18. **SemantikaLT.** Syntactic-semantic analysis and search system for Lithuanian Internet, corpus and public sector applications (2012-2014). Contracting authority: Information Society Development Committee (IVPK). Supported by Structural funds of EU, No VP2-3.1-IVPK-12-K, 2014.
  19. **Smith, B., Welty, C., McGuinness, D. L.** OWL Web Ontology Language Guide. W3C Recommendation 10 Feb 2004. Available from: <https://www.w3.org/TR/2004/REC-owl-guide-20040210/> [Accessed: 08 Apr 2016].
  20. **Son, J., Jeong, D., Baik, D.** Practical Approach: Independently Using SPARQL-to-SQL Translation Algorithms on Storage. In *Proceedings of the 4th International Conference on Networked Computing and Advanced Information Management*, Gyeongju, Korea, September 2-4, 2008, 598-603.
  21. **Wang, C., Xiong, M., Zhou, Q., Yu, Y.** Panto: A portable natural language interface to ontologies. In: *Proceedings of the 4th European conference on The Semantic Web: Research and Applications*, Innsbruck, Austria, June 3-7, 2007, 473-487.

22. **Watt, W. C.** Habitability. *American Documentation*, 19(3), July 1968, 3386-351.

## **AUTORIAUS PUBLIKACIJŲ DISERTACIJOS TEMAS RAŠTAS**

### **Mokslin s informacijos instituto (ISI) pagrindinio s rašo leidiniuose**

1. **Trakys, Algirdas ; Ablonskis, Linas; Nemurait , Lina; Paradauskas, Bronius.** A Grammar for Advanced SBVR Editor // *Information Technology and Control*, 45(1), IT&C 2015. ISSN: 1392-124X, p. 27-41.
2. **Trakys, Algirdas; Nemurait , Lina; Butkien , Rita.** SBVR based Natural Language Interface to Ontologies // *Information Technology and Control*, 46(1), IT&C 2017. ISSN: 1392-124X, p. 118-137.

### **Kituose Mokslin s informacijos instituto (ISI) duomen baz se referuojamuose leidiniuose**

1. **Trakys, Algirdas; Nemurait , Lina; Paradauskas, Bronius.** Representing and transforming SBVR question patterns into SPARQL // *Information and software technologies : 18th International Conference, ICIST 2012, Kaunas, Lithuania, September 13-14, 2012 : proceedings / [edited by] Tomas Skersys, Rimantas Butleris, Rita Butkiene.* Berlin, Heidelberg : Springer, 2012. (Communications in computer and information science, Vol. 319, ISSN 1865-0929), ISBN 9783642333071. p. 436-451. DOI: 10.1007/978-3-642-33308-8. [Conference Proceedings Citation Index]. [0,333]
2. **Trakys, Algirdas; Nemurait , Lina; Trinkevi ius, Edvinas; Paradauskas, Bronius.** Querying ontologies on the base of semantics of business vocabularies and business rules // *Information Technologies' 2011 : proceedings of the 17th international conference on Information and Software Technologies, IT 2011, Kaunas, Lithuania, April 27-29, 2011 / Edited by R. Butleris, R. Butkiene ; Kaunas University of Technology.* Kaunas : Technologija. ISSN 2029-0020. 2011, p. 247-254. [Conference Proceedings Citation Index]. [0,250]

### **Straipsniai kituose recenzuojamuose mokslo leidiniuose**

1. **Trakys, Algirdas; Nemurait , Lina; Paradauskas, Bronius; Trinkevi ius, Edvinas.** SBVR based representation of SPARQL queries and SWRL rules for analyzing semantic relations // *Bustech 2011 [elektroninis i-teklis] : the First International Conference on Business Intelligence and Technology, September 25-30, Rome, Italy. [S.l.] : IARIA, 2011. ISBN 9781612081601.* p. 1-6. [0,250].;
2. **Trakys, Algirdas; Nemurait , Lina.** Semantini uflklus vykdymas saugant ontologij reliacin je duomen baz je // *Informacin s technologijos = Proceedings of Master and PhD students conference on informatic : 15-osios tarpuniversitetin s magistrant ir doktorant konferencijos "Informacin visuomen ir universitetin s studijos" (IVUS 2010) medfliaga, 2010 m.*

gegūfl s 13 d, Kaunas, Lietuva / [Vytauto Didžiojo universitetas, Kauno technologijos universitetas, Vilniaus universitetas]. Kaunas : Vytauto Didžiojo universitetas. ISSN 2029-249X. 2011, nr. 15, p. 145-151. [0,500].;

3. <sup>T</sup>ukys, Algirdas; Nemurait , Lina; Paradauskas, Bronius; <sup>T</sup>ukėvičius, Edvinas. Transformation framework for SBVR based semantic queries in business information systems // Bustech 2012 [elektroninis žurnalas] : the second International Conference on Business Intelligence and Technology, July 22-27, 2012, Nice, France. [S.l.] : IARIA, 2012. ISBN 9781612082233. p. [1-6]. [0,250].;
4. Vileiniškis, Tomas; <sup>T</sup>ukys, Algirdas; Butkienė, Rita. An approach for semantic search over Lithuanian news website corpus // IC3K 2015 : proceedings of the 7th international joint conference on knowledge discovery, knowledge engineering and knowledge management, Vol. 1: KDIR, Lisbon, Portugal, November 12-14, 2015 Setúbal: Science and technology publications, ISBN 9789897581588. p. 57-66. [Indeksas: 0,333]  
[Indeksas grupėje: 1,333]

## INFORMACIJA APIE AUTORI

- 2010-2015. Informatikos inžinerijos mokslo krypties doktorantūros studijos KTU Informatikos fakultete.
- 2008-2010. Informatikos inžinerijos magistro studijos KTU Informatikos fakultete. gytas informatikos inžinerijos magistro laipsnis.
- 2004-2008. Bakalauro studijos KTU Informatikos fakultete. gytas informatikos bakalauro kvalifikacinis laipsnis.
- Moksliniai interesai srityse: veiklos flodynė ir veiklos taisyklių semantika, ontologijos, ontologijų sklaidos kalbos, semantinių technologijų, duomenų bazės, informacinių sistemų projektavimas ir dalykinės sritys modeliavimas.
- El. paštas: algirdas.sukys@ktu.lt

## **ABSTRACT**

### **QUERYING ONTOLOGIES ON THE BASE OF SEMANTICS OF BUSINESS VOCABULARY AND BUSINESS RULES**

The amount of information on the Web grows constantly nowadays. Information overload makes a Web search process tedious. Traditional keyword based search engines analyse HTML documents that are intended to render information for humans but does not represent semantics, which a computer can understand. Even though such search engines help to find information, they give redundant or incomplete results based on keyword matches, leaving a lot of work for users to find relevant information. For example, it would be a difficult task to find all heads of the European Union states using keyword based search. The user would have to put additional effort into completing this search.

The Semantic Web idea [1] is based on understanding the meaning of published information and processing it by machines. The backbone of a Semantic Web is ontologies that store entities, representing real world objects (i.e., persons, vehicles, organizations), their relations, properties, etc. The search across ontologies is called semantic search. Due to the capability to understand the intent of the user's queries and even complex questions, semantic search returns results that are more precise.

One of the challenges of developing a system with a semantic search function is the implementation of a usable and convenient user interface. A number of interfaces to ontologies were introduced after the Semantic Web idea spread: Semantic Crystal [11], Ginseng [1], QuestIO [6], FREyA [4], ORAKEL [3], PANTO [21], Querix [12], etc. They vary from simple interfaces for SPARQL queries to more sophisticated natural language interfaces (NLIs) and differ in their usability. The study of E. Kaufmann and A. Bernstein [11] was carried out to compare keyword-based search, graphical query language, natural language and menu guided interfaces. It was found, that users prefer querying ontologies using full sentences in natural language. The research revealed the potential of NLIs for end-user access to the Semantic Web, as this type of interface proved the most useful and best-liked query interface.

#### **Motivation**

In this work, it was decided to create a new NLI. The first reason was the desire to write questions in multiple languages. It is important, because 25.9% of internet users use the English language, while the other users use other languages [15]. Existing NLIs show good results answering questions in English. However, authors do not discuss about adapting their solutions for other languages, i.e., which components are independent from language, and which should be replaced or adjusted, what source code modifications are required. Certainly, simple replacement of standard linguistic libraries (e.g., Stanford parser, WordNet, etc.)

would not be enough, it would require a significant source code modification.

Another reason is about mapping questions with ontology resources (i.e., classes, properties, etc.). This is a critical function of NLI, required for translating questions to queries. Ontologies in the Semantic Web are processed and understood by machines. The problem is that their structure can differ from how people think about data and formulate questions. It is obvious, that people desire writing simple questions, while data in ontologies can be stored using complex structures (e.g., using n-ary relations). Therefore, straightforward mappings (i.e., question to a single ontology resource) is not enough, NLI must be able to perform complex mappings (i.e., question to a combination of resources). The analysis of existing NLIs to ontologies revealed that most of them extract lexicon directly from ontologies. As a result, they allow only straightforward mapping and understand only those questions that correspond to the structure of the ontology.

Therefore, the basic principles of the system's architecture was formulated: NLI must be adjustable for different languages and the lexicon must allow relating complex ontological constructions with simple questions. It was decided to use SBVR in order to achieve this. This standard is intended to specify business vocabularies and business rules using structured natural language. The foundation of SBVR is a semiotic/semantic triangle, which is the theoretical basis for SBVR's linguistic based architecture that separates expression from meaning [17]. It allows the expressing of the same things differently as well as in different languages. Therefore, a question, written in different languages, has the same model of meaning and can be transformed to an ontology query regardless of the language it is written in.

SBVR vocabulary concepts can have definitions given as rules that describe derivations of those concepts. Such definitions formally specify the derivation of concepts from other concepts and can support inferences [13]. This suggests that SBVR definitions could be used to bridge the gap between the way in which a particular item of data is stored (i.e., the ontology scheme) and the way of, how a user thinks about the data and formulates questions.

Although the SBVR metamodel supports questions and allows querying software models, it was not previously used for semantic search. This work should answer, whether or not SBVR can be used as a basis for NLI, which is multilingual and allows mapping simple questions with complex ontology structures (i.e., combinations of ontology resources).

### **Object and scope of research**

The object of this research is a process of querying ontologies using natural language questions. The scope of the research includes the following topics:

- Natural language interfaces to knowledge bases;
- Most advanced knowledge and data models, their representation and query languages (SBVR, OWL 2, RDF, SPARQL), related tools and technologies;

- Model driven transformation technologies.

### **Problem statement and research questions**

The lack of usable and convenient user interfaces to ontologies, allowing questioning in natural languages is the problem inspired by this research. When solving this problem it was important to fulfil such requirements:

- Adjustability to questioning in different languages (i.e., languages, investigated in this work: English and Lithuanian; and grammatically similar languages: German, Czech, Polish, etc.);
- Ability to map questions with combinations of ontology resources;
- Portability (i.e., ability to question in different domains);
- Effectiveness of answering questions similar to other NLI.

This research intends to answer the following questions:

1. Is it possible to use SBVR questions for querying ontologies and relating natural language questions with combinations of ontology resources?
2. How natural language questions can be transformed to SPARQL using SBVR?
3. Is it possible to achieve portability without compromising the correctness of NLI to ontologies using SBVR?
4. Can SBVR based NLI to ontologies be adjusted to different languages and what components are language specific?

### **Goals and objectives**

The main goal of this work is to extend semantic search capabilities, allowing users to write natural language questions in different languages, also including such cases when mapping of questions to ontology is complex (i.e., questions must be mapped with combinations of ontology resources). Research tasks are the following:

1. To analyse literature related with OWL ontologies and ontology query language SPARQL; existing NLI to databases and ontologies; SBVR knowledge model and tools to write SBVR specifications;
2. To define the conception of NLI to ontologies and algorithms for transforming natural language questions to SPARQL queries;
3. To define the conception of a SBVR tool for writing business vocabularies, rules, and questions;
4. To create prototypes for evaluating the relevance of the solution;
5. To conduct an experiment and evaluate research results.

The scope of this work can solve just a limited set of problems, related with natural language questions, sufficient for proving the concept. The main quality criteria for the solution are as follows:

- Ability to question ontologies in different languages;
- Ability to map questions with combinations of ontology resources;



- Portability of the solution;
- Effectiveness of answering questions.

### Research methodology

The research was carried out using the methodology of Design Science (also called constructive) research. This paradigm seeks to extend the boundaries of human and organizational capabilities by creating new and innovative artefact [10]. This artefact can be algorithm, framework, model, etc. In this research, it is Semantic search solution. The methodology and steps of the research are presented in Table 1.

**Table 1.** The research methodology

<p>1. Selecting a practically relevant research problem.</p>	<p>It was found that the process of traditional keyword based search leaves a lot of manual work for users to find relevant information from results of keyword matching.</p>
<p>2. Analysing existing solutions to find out the potential for the research.</p>	<p>The search process can be facilitated performing semantic search over ontologies. The most convenient interface for ontologies is NLI. The analysis of existing NLIs to ontologies showed that existing solutions have limitations. Therefore, it was decided to create a new Semantic search solution.</p>
<p>3. Analysing the domain to understand the problem and create the solution to solve it.</p>	<p>First, ontologies and ontology query languages were analysed to understand, how ontologies are modelled and queried. To understand the area of creating NLIs (e.g., problems that are faced creating and using such systems, their main features, requirements, etc.), research, related with NLIs to databases and ontologies were analysed. Finally, having an insight to solve the shortcomings of existing NLIs, SBVR standard and capabilities to use it for Semantic search solution were analysed.</p>

4. Creating the original solution	The conception of SBVR based Semantic search solution was created. The solution consist of NLI to ontologies and SBVR structured language editor (SBVR SLE). The solution was theoretically described by defining rules to transform natural language questions to SPARQL queries and grammar for creating SBVR structured language editor.
5. Implementing the prototype of the solution and evaluating it	Two prototypes were implemented: SBVR SLE and NLI to ontologies. Experiments were conducted for evaluating the applicability of the solution and comparing it against other similar solutions.
6. Feasibility to apply the solution in practice	The created solution can be applied to implement the semantic search in the Web. It is also expected that the created SBVR SLE will create conditions for other SBVR related research.
7. Relations of the solution with theoretial studies	The created solution will complement the set of available solutions of NLIs to ontologies. This work also contributes to the research of SBVR and presents the applicability of this standard for querying.

### **Defended propositions**

Propositions defended by this thesis are the following:

1. SBVR allows using language independent rules to transform questions to semantic queries and achieve multilingualism of NLI to ontologies. However, such a solution also requires language dependent components to perform syntactic and morphological analysis of questions.
2. SBVR derivation rules allows describing the relations between natural language questions and complex ontology structure. As a result, ontologies can be queried written questions that do not directly correspond to their structure.

3. SBVR based NLI is portable. Portability is achieved specifying business vocabulary and rules (i.e. lexicon) of certain domain and mapping it with ontology. Mappings are performed labelling ontology resources with representations of corresponding vocabulary concepts.

### **Major contributions and novelty**

The major contribution of this work is the solution of querying ontologies using natural language. The solution consists of the following two parts:

- NLI to ontologies, which allows writing, analysing natural language questions and transforming them to SPARQL queries.
- The SBVR SLE that is used for configuration of NLI. It allows specifying business vocabularies, business rules and writing questions using structured language. In addition, specifications can be transformed to the SBVR XMI model.

The novelty of the research is as follows:

1. It was not found in any research, published by other authors, the using SBVR questions for querying ontologies.
2. The created NLI to ontologies has clearly defined parts, which have to be replaced for querying in different languages.
3. The created NLI to ontologies allows questioning when the structure of ontologies is complex and it is needed to map questions with combinations of ontology resources. These mapping are defined using SBVR derivation rules or formal definitions of SBVR concepts.
4. The created SBVR SLE is used for writing specifications of business vocabularies and business rules to formulate questions. This tool allows the splitting of specifications into separate parts and hide some vocabulary entries (i.e., those that are used only for derivations) from user. In addition, this feature allows the creating of metavocabularies and applying editor for other purposes (e.g., transformation to OWL 2, specification of BPMN business processes, etc.).

### **Practical significance**

The presented solution offers a way to implement semantic search writing questions in natural language. It can be configured to questions for ontologies of different domains and in different languages. The ability to define SBVR derivations allows the answering of simple questions in complex ontologies that are often used in practice. The solution can be applied for a semantic search on the Web or in business applications.

The results of this work were applied in the SemantikaLT project [18], creating a semantic search service that allows writing questions in the Lithuanian language. The SBVR SLE was used for creating SBVR business vocabularies, business rules and configuration of NLI. The transformations, created in this work,

were used to transform Lithuanian questions to SPARQL for querying semantically annotated Lithuanian Internet corpora for Politics, Business and Economy, and Public Administration domains.

The created SBVR SLE can be used to other research, related with SBVR to create specifications of business vocabulary and business rules.

## Conclusions

1. Analysis of scientific publications has shown that the preferable interface for querying ontologies is a natural language interface (NLI). The most important requirements of such a NLI are adjustability to different languages, ability to deal with complex structures of ontology resources, portability and habitability. Existing NLIs to ontologies only allow questioning in English and usually in the form when the formulation of question directly corresponds to the structure of ontology. On the other hand, existing solutions are portable and can be configured for questioning in different domains. To improve habitability, NLIs use various techniques that can be divided into two groups: (1) methods intended to familiarize users with lexicon and help formulating questions; (2) methods intended to help interpreting and disambiguating questions.
2. The analysis of SBVR standard has shown that a distinguishing feature of its metamodel to separate the meaning from representation allows achieving multilingualism. It suggests that questions in the SBVR based NLI could be transformed to semantic queries using language-independent rules, because the model of a question's meaning is the same for all languages. The language dependent components of such a NLI should only be those that help writing and interpreting questions. In order to achieve multilingualism, these components should be replaced or adjusted for questioning in a certain language. The architecture of the implemented NLI was designed in pursuance of these ideas. Another important aspect related with deciding to use SBVR for NLI to ontologies is the derivation rules that can be used to relate simple questions with complex ontology structure.
3. The analysis of SPARQL revealed that the syntax of this query language and metamodel of SBVR questions has conforming elements, expressing information needs and restrictions of a query. It led to the assumption that transformation of SBVR questions to SPARQL is feasible. Therefore, it was decided to describe detailed mappings between the metamodel of SBVR questions and SPARQL and create transformation rules.
4. An important prerequisite to use SBVR for NLI to ontologies is a robust SBVR editing tool for specifying vocabulary and rules (i.e., configuration of NLI) and generating XMI models of questions for further transformations. The analysis of existing tools showed that none of them meets the

requirements and their further improvements are complicated. Therefore, it was decided to create a new SBVR editor.

5. The conception of SBVR editor is based on structured language grammar (in EBNF-like form) for specifying business vocabularies, rules, and writing questions. The grammar was described analysing the metamodel of SBVR representations, structured language examples from SBVR specification and practice. Grammar supports questions to retrieve; objects of a certain type, questions with modifier attachments, cardinality restrictions, numeric comparisons, and count function.
6. Xtext framework was used to describe the grammar and implement the SBVR editor. Experiments have shown that the editor allows specifying all the required SBVR constructions. It also creates the necessary preconditions for portability and a multilingual NLI to ontologies, along with being suitable for the configuration task.
7. The conception of SBVR based NLI to ontologies contains the following components (only the first two are language-dependent):
  - User interface, which allows formulating natural language questions, conforming to SBVR vocabulary;
  - Question analyser, which identifies SBVR concept (s) that the question is based on;
  - SBVR model composer, which constructs the question's SBVR XMI model;
  - Component of query transformation, which transforms the question's SBVR XMI model to SPARQL query.
8. Implementation of SBVR based NLI proved that it is possible to use SBVR questions for querying ontologies. Additionally, experimental investigation shows that the solution allows querying ontologies, whose structure (i.e., expressing relevant part of domain knowledge) directly does not correspond to the structure of the natural language questions. This is achieved specifying derivation rules of SBVR concepts that are used when transforming questions to SPARQL.
9. Experiments of questioning in two different domains proved that it is possible to achieve the portability and multilingualism of a NLI that uses SBVR standard. Portability is achieved by allowing SBVR specification to be written for a certain domain and linking it with the ontology. The evaluation of effectiveness showed similar result as other NLIs (i.e. f-measure is 0,86 in the domain of geography and 0,75 in the domain of restaurants). However, the main advantages of the created solution is multilingualism and the ability to question ontologies, whose structure does not directly correspond to the structure of natural language questions.

UDK 004.775 (043.3)

SL344. 2017-05-09, 2 leidyb. apsk. I. Tiražas 50 egz.

I-leido Kauno technologijos universitetas, K. Donelaičio g. 73, 44249 Kaunas  
Spausdino leidyklos „Technologija“ spaustuv., Studentų g. 54, 51424 Kaunas