

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
INFORMACINIŲ SISTEMŲ INŽINERIJOS STUDIJŲ PROGRAMA

INGRIDA GREGERYTĖ

SEMANTINIO ANOTAVIMO METODIKA SPORTO
ĮVYKIAMS ATPAŽINTI LIETUVIŠKAME TEKSTE

Magistro darbas

Darbo vadovė
doc. dr. R. Butkienė

KAUNAS, 2013

KAUNO TECHNOLOGIJOS UNIVERSITETAS
INFORMATIKOS FAKULTETAS
INFORMACINIŲ SISTEMŲ INŽINERIJOS STUDIJŲ PROGRAMA

INGRIDA GREGERYTĖ

SEMANTINIO ANOTAVIMO METODIKA SPORTO
ĮVYKIAMS ATPAŽINTI LIETUVIŠKAME TEKSTE

Magistro darbas

Darbo vadovė:
doc. dr. R. Butkienė
2013-05-

Recenzentas:
doc. dr. L. Motiejūnas
2013-05-

Atliko:
IFM-1/7 gr. studentė
Ingrida Gregerytė
2013-05-

KAUNAS, 2013

AUTORIŲ GARANTINIS RAŠTAS DĖL PATEIKIAMO KŪRINIO

2013 - Gegužės - 23 d.
Kaunas

Autorius, _____ Ingrida Gregerytė _____,
(vardas, pavardė)

patvirtina, kad Kauno technologijos universitetui pateiktas baigiamasis bakalauro (magistro) darbas (toliau vadinama – Kūrinys) _____ „Semantinio anotavimo metodika sporto įvykiams atpažinti
(kūrinio pavadinimas)
lietuviškame tekste“ _____

pagal Lietuvos Respublikos autorių ir gretutinių teisių įstatymą yra originalus ir užtikrina, kad

- 1) jį sukūrė ir parašė Kūrinyje įvardyti autoriai;
- 2) Kūrinys nėra ir nebus įteiktas kitoms institucijoms (universitetams) (tiek lietuvių, tiek užsienio kalba);
- 3) Kūrinyje nėra teiginių, neatitinkančių tikrovės, ar medžiagos, kuri galėtų pažeisti kito fizinio ar juridinio asmens intelektinės nuosavybės teises, leidėjų bei finansuotojų reikalavimus ir sąlygas;
- 4) visi Kūrinyje naudojami šaltiniai yra cituojami (su nuoroda į pirminį šaltinį ir autorių);
- 5) neprieštarauja dėl Kūrinio platinimo visomis oficialiomis sklaidos priemonėmis.
- 6) atlygins Kauno technologijos universitetui ir tretiesiems asmenims žalą ir nuostolius, atsiradusius dėl pažeidimų, susijusių su aukščiau išvardintų Autorių garantijų nesilaikymu;
- 7) Autoriai už šiame rašte pateiktos informacijos teisingumą atsako Lietuvos Respublikos įstatymų nustatyta tvarka.

Autorius

(vardas, pavardė)

(parašas)

SUMMARY

Methodology of Semantic Annotation for Recognition of Sport Events in the Lithuanian Text

Worldwide technology advances aimed at creating a user friendly semantic network. Also facilitate human-software interaction, ie to create conditions for computer to understand natural human language. One of the steps in order to achieve this goal is the automatic semantic annotation.

Currently, most of the information is on the website. With such a large amount of information, it becomes difficult to find exactly what you need. The search in semantically annotated texts can obtain more accurate results than using a simple web search. However, in order to semantically annotate the text, specialized dictionaries of domains and ontologies have to be created, suitable annotation methods chosen, semantic annotation rules described and tools for that created. It's especially important for Lithuanian language to develop tools of semantic annotation of text because it is behind in this area compared to other languages.

The main purpose of this work is to create a methodology to recognize sport events in the Lithuanian text and provide an ontology which allows to annotate such text. According described methodology it is possible to recognize achievements defeats, results and participants of event. Created methodology and described sport events rules can be used to create an automatic semantic annotation tool.

Ontology was created in OWL language, using ontology development tool " Protégé ". Because tool is free and chosen language is widely used, everyone can use and modify this ontology. Sporting events ontology can be supplemented new classes or other ontologies. SPARQL queries were developed, which are used to perform searches in sporting events ontology.

SANTRAUKA

Semantinio anotavimo metodika sporto įvykiams atpažinti lietuviškame tekste

Pasaulyje, tobulėjant technologijoms, siekiama sukurti draugiškesnį vartotojams semantinį tinklą. Taip pat palengvinti žmogaus ir programinės įrangos bendravimą, t.y. sukurti prielaidas, kompiuteriui suprasti natūralią žmonių kalbą. Vienas iš žingsnių, norint įgyvendinti šį siekį, yra automatinis semantinis teksto anotavimas.

Šiuo metu daugiausia informacijos yra pateikiama internetiniuose puslapiuose. Esant tokiam dideliame informacijos kiekiui, darosi sunku surasti būtent tai ko reikia. Atliekant paiešką semantiškai anotuotuose tekstuose galima gauti tikslesnius rezultatus, negu naudojant paprastas paieškos sistemas. Tačiau norint semantiškai anotuoti tekstą, visų pirma turi būti sukurti specializuoti sričių žodynai, ontologijos, pasirinkti tinkami anotavimo metodai, aprašytos semantinio anotavimo taisyklės, sukurti tam skirti įrankiai. Lietuvių kalbai ypač svarbu sukurti priemones semantinio teksto anotavimui, nes ji šioje srityje atsilieka lyginant su kitomis kalbomis.

Pagrindinis šio darbo tikslas buvo sukurti metodiką, skirtą atpažinti sporto įvykius lietuviškame tekste ir pateikti ontologiją, kuria naudojantis būtų galima anotuoti tokį tekstą. Pagal aprašytą sporto įvykių atpažinimo metodiką tekste galima atpažinti laimėjimus ir pralaimėjimus, rezultatus, įvykyje dalyvavusius dalyvius. Sukurtą metodiką ir aprašytas sporto įvykių taisykles galima panaudoti, kuriant automatinio semantinio anotavimo įrankį. Metodiką galima patobulinti, kad būtų galima atpažinti įvykius esančius ne tik sakinyje, o visame tekste.

Ontologija buvo sukurta OWL kalba, naudojantis ontologijos kūrimo įrankiu Protégé. Kadangi šis įrankis yra nemokamas, o pasirinkta kalba yra plačiai naudojama, naudotis ontologija ir ją tobulinti gali visi norintys. Sporto įvykių ontologiją galima papildyti naujomis klasėmis, kitomis ontologijomis. Sudarytos SPARQL užklauskos, kuriomis naudojantis sporto įvykių ontologijoje atliekama paieška.

Turinys

LENTELIŲ SĄRAŠAS	7
PAVEIKSLŲ SĄRAŠAS.....	8
TERMINŲ IR SANTRUMPŲ ŽODYNAS.....	9
ĮVADAS.....	10
1. PROBLEMINĖS SRITIES ANALIZĖ	12
1.1. ANALIZĖS TIKSLAS	12
1.2. TYRIMO OBJEKTAS, SRITIS IR PROBLEMA.....	12
1.3. TYRIMO OBJEKTO ANALIZĖ.....	13
1.4. VARTOTOJŲ ANALIZĖ	15
1.4.1. <i>Vartotojų aibė, tipai ir savybės</i>	15
1.4.2. <i>Vartotojų tikslai ir problemos</i>	15
1.5. ESAMŲ SPRENDIMŲ, SKIRTŲ SEMANTINIAM ANOTAVIMUI, ANALIZĖ.....	16
1.5.1. <i>KIM</i>	17
1.5.2. <i>MnM</i>	18
1.5.3. <i>GATE</i>	18
1.5.4. <i>OntoMat Annotizer</i>	19
1.5.5. <i>Semantinio anotavimo platformų palyginimas</i>	20
1.6. ESAMŲ ONTOLOGIJŲ KŪRIMO ĮRANKIŲ ANALIZĖ	20
1.6.1. <i>Protégé</i>	20
1.6.2. <i>TopBraid Composer</i>	21
1.6.3. <i>Ontologijų kūrimo įrankių palyginimas</i>	22
1.7. ESAMŲ SPORTO IR ĮVYKIŲ SRIČIŲ ONTOLOGIJŲ ANALIZĖ	23
1.7.1. <i>BBC Sporto ontologija</i>	23
1.7.2. <i>Bendroji įvykių ontologija</i>	25
1.7.3. <i>Esamų ontologijų panaudojimas</i>	27
1.8. ONTOLOGIJOS KŪRIMO METODIKA	27
1.9. DARBO TIKSLAS IR UŽDAVINIAI.....	28
1.10. SIEKIAMO SPRENDIMO APIBRĖŽIMAS	28
1.11. ANALIZĖS IŠVADOS	28
2. SPRENDIMO REIKALAVIMŲ SPECIFIKACIJA.....	30
2.1. REIKALAVIMŲ METODIKAI SPECIFIKACIJA.....	30
3. SPORTO ĮVYKIŲ ONTOLOGIJOS REALIZACIJA.....	32
3.1. PAGRINDINĖS ONTOLOGIJOS KLASĖS	32
3.2. ONTOLOGIJOS OBJEKTŲ SAVYBĖS.....	34
3.3. DUOMENŲ TIPŲ SAVYBĖS	35
4. ĮVYKIO ATPAŽINIMO TEKSTE METODIKA	37
4.1. ĮVYKIO ATPAŽINIMO TAISYKLĖS.....	37
4.2. ĮVYKIO ATPAŽINIMO METODIKOS TAIKYMAS	51
5. EKSPERIMENTINIS TYRIMAS	53
5.1. EKSPERIMENTO PLANAS	53
5.2. EKSPERIMENTO VYKDYMAS	53
5.3. EKSPERIMENTO REZULTATŲ ANALIZĖ	59
5.4. SPRENDIMO TAIKYMO REKOMENDACIJOS	59
6. REZULTATŲ APIBENDRINIMAS IR IŠVADOS	60
7. LITERATŪRA	61
8. PRIEDAI.....	64
8.1. PRIEDAS. TEKSTŲ FRAGMENTAI.....	64
8.2. PRIEDAS. TEKSTŲ MORFOLOGINIO ANOTAVIMO REZULTATAS	64
8.3. PRIEDAS. TEKSTŲ SINTAKSINIO ANOTAVIMO REZULTATAS	71
8.4. PRIEDAS. TEKSTŲ SEMANTINIO ANOTAVIMO PAVYZDŽIAI	71

LENTELIŲ SĄRAŠAS

1.1 lentelė. Vartotojai.....	15
1.2 lentelė. Vartotojų tikslai ir problemos.....	16
1.3 lentelė. Semantinio anotavimo platformų palyginimas.....	20
1.4 lentelė. Ontologijų kūrimo įrankių palyginimas	22
4.1 lentelė. Šablonai skirti komandos atpažinimui	38
4.2 lentelė. Šablonai skirti dalyvio atpažinimui	40
4.3 lentelė. Šablonai skirti sporto renginio atpažinimui.....	41
4.4 lentelė. Šablonai skirti rezultato atpažinimui.....	47
4.5 lentelė. Šablonai skirti kėlinio rezultato atpažinimui.....	49
8.1 lentelė. Eksperimento tekstų fragmentai	64
8.2 lentelė. Morfologiškai anotuoti tekstai.....	64
8.3 lentelė. Sintaksiškai anotuotas tekstas	71
8.4 lentelė. 2 tekstas, pažymėtos pagrindinės anotacijos.	71
8.5 lentelė. 3 tekstas, pažymėtos pagrindinės anotacijos.	72
8.6 lentelė. 4 tekstas, pažymėtos pagrindinės anotacijos.	73

PAVEIKSLŲ SĄRAŠAS

1.1 pav. Sporto įvykių dalykinės srities esybių modelis	13
1.2 pav. Semantinis anotavimas [7]	13
1.3 pav. Bendras semantinio anotavimo procesas	14
1.4 pav. Bendra semantinio anotavimo proceso įeiga ir išeiga	15
1.5 pav. Bendroji semantinio anotavimo platformų architektūra [14]	16
1.6 pav. Semantinio anotavimo platformų klasifikacija [12]	17
1.7 pav. Semantinis anotavimas naudojantis KMI [13]	18
1.8 pav. Komponentų modelis [16]	19
1.9 pav. Protégė langas	21
1.10 pav. TopBraid Composer Free Editions langas	22
1.11 pav. Olimpiados 2012 dviračių rungtynių modelis [18]	24
1.12 pav. Įvykių ontologijos modelis [19]	26
2.1 pav. Automatinis sporto įvykių atpažinimo procesas	31
3.1 pav. Sporto įvykių ontologija	33
3.2 pav. Sporto įvykių ontologijos objektų savybės	35
3.3 pav. Sporto įvykių ontologijos duomenų tipų savybės	36
4.2 pav. Komandos atpažinimo procesas	38
4.3 pav. Sporto renginio atpažinimo procesas	41
4.4 pav. Varžybų dalyvio atpažinimo procesas	43
4.5 pav. Laimėjimo/ pralaimėjimo atpažinimas	44
4.6 pav. Varžybų rezultato atpažinimo procesas	47
4.7 pav. Varžybų kėlinio rezultato atpažinimo procesas	48
4.8 pav. Dalyvio rezultato atpažinimo procesas	50
4.9 pav. Įvykio atpažinimo metodikos taikymo procesas	52
5.1 pav. Užklauskos „Kas laimėjo rungtynes tarp komandos A ir komandos B?“ rezultatas	55
5.2 pav. Užklauskos „Kokios komandos kovojo mače?“ rezultatas	56
5.3 pav. Užklauskos „Kokios komandos kovojo mače X?“ rezultatas	56
5.4 pav. Užklauskos „Kokiu rezultatu laimėjo komanda A prieš komandą B?“ rezultatas	56
5.5 pav. Užklauskos „Koks teisėjas teisėjavo rungtynėse X?“ rezultatas	Error! Bookmark not defined.
5.10 pav. Užklauskos „Kada įvyko įvykis?“ rezultatas	57
5.6 pav. Užklauskos „Kas laimėjo pasaulio futbolo čempionatą?“ rezultatas	57
5.7 pav. Užklauskos „Kas įmušė įvartį per varžybas X?“ rezultatas	57
5.8 pav. Užklauskos „Kiek žaidėjas Y įmetė taškų per rungtynes X?“ rezultatas	58
5.9 pav. Užklauskos „Kokį rekordą pasiekė sportininkė Y per čempionatą X?“ rezultatas	58
5.11 pav. Užklauskos „Kokį greitį pasiekė atletas rungtynėse X?“ rezultatas	58

TERMINŲ IR SANTRUMPŲ ŽODYNAS

Semantinis anotavimas - teksto pažymėjimas ontologijos konceptu ir įtraukimas į ontologiją[7];

NER (*Named Entity Recognition*) – įvardintų esybių atpažinimas;

POS (*Part Of Speech*) - morfologinis anotavimas, kurio metu teksto dalys pažymimi gramatinėmis pažymomis;

Sintaksinis anotavimas – sakinio dalių pažymėjimas;

Ontologija – dalykinės srities terminų ir santykių tarp jų specifikacija[4];

Tekstynas – bet koks, kad ir pats mažiausias, elektroninę formą turinčių tekstų rinkinys [45];

Anotuotas tekstynas – tekstynas papildytas aiškinamąja kalbine informacija;

KIM (*Knowledge and Information Management*) – semantinio anotavimo platforma;

API (*Application Programming Interface*) - aplikacijų programavimo sąsaja;

OKBC (*Open Knowledge Base Connectivity*) – protokolas, numatantis rinkinių operacijų, skirtų bendrai vartotojo sąsajai žiniomis grindžiamoms sistemoms[42];

OWL (*Web Ontology Language*) – ontologijų kūrimo kalba;

IDE (*Integrated Development Environment*) – integruota kūrimo aplinka;

RDF (*Resource Description Framework*) – išteklių aprašymo sistema;

SPARQL – užklausų kalba, skirta RDF;

Taksonomija – klasifikacija, kurioje elementai išdėstyti medžio struktūros principu;

JAPE (*Java Annotation Patterns Engine*) – Java anotacijos modelio variklis, skirtas atpažinti reguliarias išraiškas tekste.

ĮVADAS

Darbas parengtas pagal Informacinių sistemų inžinerijos magistrantūros studijų programos reikalavimus.

Darbo problematika ir aktualumas

Šiuo metu didžiausias informacijos kiekis yra pateikiamas internetiniuose puslapiuose. Kiekvieną dieną atsinaujinančią informaciją teikia naujienų portalai. Tokį kiekį informacijos yra sudėtinga, o kartais ir neįmanoma peržiūrėti, bei surasti reikiamą informaciją. Šiuo atžvilgiu, semantiškai anotuotuose tekstuose rasti reikiamą informaciją būtų daug paprasčiau ir patogiau. Anotavimo sprendimai lietuvių kalbai, kitų kalbų atžvilgiu, šioje srityje smarkiai atsilieka. Todėl, norint, kad lietuvių kalba būtų ir toliau naudojama pasauliniame tinkle, reikia sukurti įrankius skirtus lietuviško teksto semantiniam anotavimui. Tokie sprendimai palieštų ne tik savo sričių specialistus (IT, kompiuterinės lingvistikos, Lietuvių kalbos), bet ir paprastus interneto vartotojus. Taip palengvindami jų darbą, reikalingos informacijos suradimą.

Lietuviškuose naujienų portaluose dažniausia naujienos būna suskirstytos į tokias temas: verslas, politika, kriminalai, sportas. Pasirinkta sporto įvykių tematika, todėl, kad tokios naujienos nuolat atsinaujina, yra aktualios.

Darbo tikslas ir uždaviniai

Šio magistro darbo tikslas yra sudaryti sąlygas semantiškai anotuoti su sporto įvykiais susijusius lietuviškus tekstus ir tokiu būdu prisidėti prie lietuvių kalbos įsitvirtinimo pasauliniame semantiniame tinkle

Uždaviniai:

1. Išanalizuoti semantinio anotavimo procesą;
2. Išanalizuoti teksto anotavimo įrankius, formatus ir standartus;
3. Sukurti sporto įvykių ontologiją;
4. Specifikuoti sporto įvykių atpažinimo tekste taisykles;
5. Parengti metodiką, skirtą atpažinti sporto įvykius lietuviškame tekste;
6. Ištirti sukurtą sporto įvykių ontologiją ir metodiką.

Darbo rezultatai ir jų svarba

Sukurta sporto įvykių atpažinimo tekste metodika ir sporto įvykių ontologija, gali būti panaudota kuriant lietuviškų tekstų automatinį semantinio anotavimo įrankį, skirtą lietuviško teksto anotavimui. Norint, kad sporto įvykiai būtų aptinkami ne tik sakinyje, bet visame tekste, aprašytas taisykles galima papildyti naujomis. Taip pat šios metodikos principus galima panaudoti kuriant metodikas kitokio tipo įvykių atpažinimui.

Darbo struktūra

Darbe yra 8 skyriai. Pirmame skyriuje pateikta tyrimo objekto, srities ir problemos, vartotojų, esamų semantinio anotavimo platformų ir ontologijų kūrimo įrankių, jau sukurtų ontologijų ir jų kūrimo metodikos analizė.

Antrame skyriuje apibrėžtas darbo tikslas, uždaviniai ir rezultato kokybės kriterijai.

Trečiame skyriuje aprašyta įvykio atpažinimo tekste metodika, metodikos taikymas ir teksto anotavimas ją panaudojant.

Ketvirtame skyriuje aprašyta sukurta sporto įvykių ontologija.

Penktame skyriuje pateikiamas eksperimentas, kurio metu sporto įvykių ontologijoje buvo atlikta paieška pagal ankščiau užduotus klausimus.

Šeštame skyriuje pateikiamos galutinės darbo išvados.

Septintame skyriuje pateikiamas naudotos literatūros sąrašas, o aštuntame - priedai.

1. PROBLEMINĖS SRITIES ANALIZĖ

1.1. Analizės tikslas

Analizės tikslas yra geriau suprasti tiriamąją sritį, objektą, išanalizuoti esamus sprendimus, jų pritaikymą lietuvių kalbai ir pagrįsti lietuviško teksto automatinio semantinio anotavimo priemonių poreikį.

1.2. Tyrimo objektas, sritis ir problema

Problema

Kitoms kalboms yra sukurti anotuoti tekstynai [22, 23], įvairūs automatinio anotavimo įrankiai: KIm, GATE. Daugiausia įdirbio ir sukurtų sprendimų yra įvardintų esybių atpažinimo (*angl. Named Entity Recognition*) srityje. Įvardintų esybių atpažinimas (NER), tai procesas, kurio metu tekste randami paminėti specifiniai dalykai. Pvz. naujienų esybės: Žmonės, vietos ir organizacijos [25]. Tai palengvina semantinį anotavimą, nes atpažinus įvardintas esybes, sužinoma kokios sričiai esybė ir priklauso ir ką ji reiškia. Automatiniam įvardintų esybių atpažinimui sukurti įrankiai arba įrankiai, kuriuose yra NER funkcija: Stanford NER [26], LingPipe [25], ANNIE [27], brat [28]. Paskutinis įrankis gali būti panaudotas ir įvykių anotavimui.

Lietuvių kalba skirti sprendimai, automatinio anotavimo srityje, atsilieka. Kol kas yra sukurta tik morfologinio anotavimo įrankis ir dalinai morfologiškai anotuotas tekstynas [46].

Taip pat semantinis anotavimas praverstų kuriant paieškos sistemas, kurios ieškotų informacijos ne pagal statistiką (taip daro „Google“ ar „Yahoo“ paieškos sistemos), o pagal prasmę. Tokiu būdu galima išgauti tikslesnius rezultatus.

Siekiant sukurti sąlygas automatiniam lietuviško teksto semantiniam anotavimui, pirmiausia reikia išnagrinėti esamus anotavimo įrankius bei metodus ir jų kūrimo patirtį pritaikyti lietuvių kalbai.

Nekreipiant į šią problemą dėmesio, gali nukentėti lietuvių kalbos vartojimas pasauliniame internetiniame tinkle, nes kitomis kalbomis bus paprasčiau gauti tikslesnę informaciją.

Tyrimo objektas

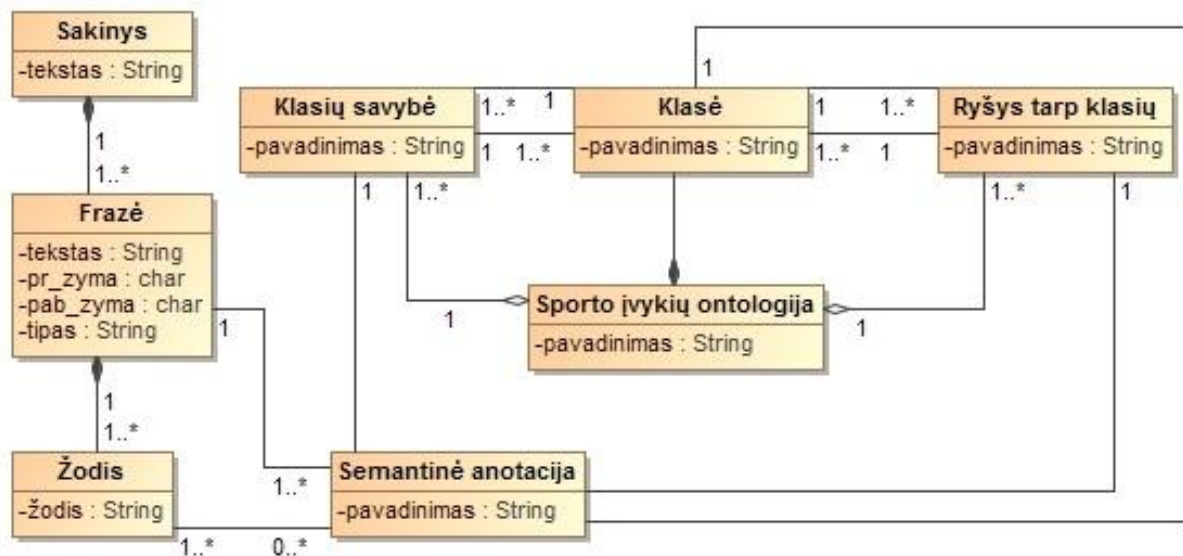
Tyrimo objektas – sporto įvykių atpažinimas lietuviškame tekste ir šių tekstų semantinis anotavimas.

Tyrimo sritis

Tyrimo sritys - semantinio anotavimo proceso automatizavimas, esami semantinio anotavimo įrankiai, formatai ir standartai.

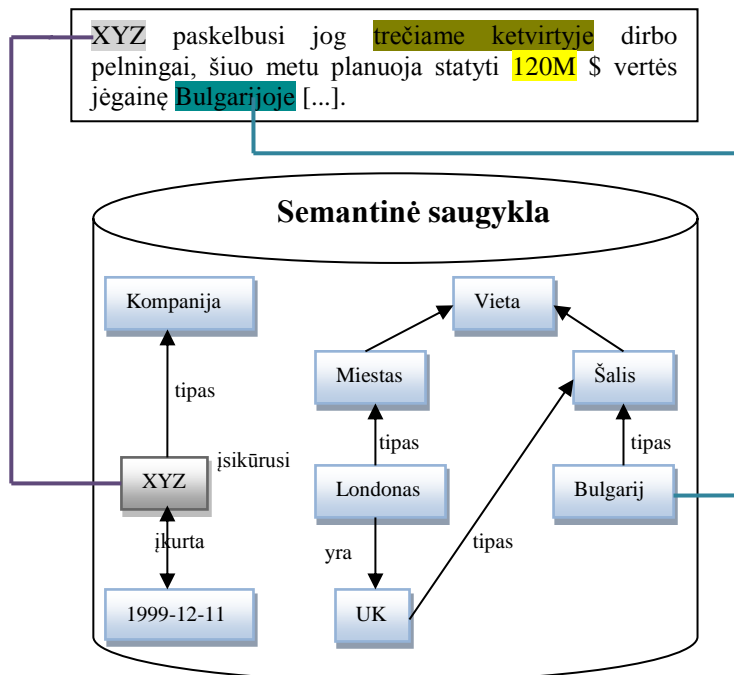
1.3. Tyrimo objekto analizė

1.1 paveiksle pateikiamas dalykinės srities esybių modelis. Pagrindinės esybės yra šios: semantinė anotacija, frazė, sakinys, žodis, ontologija, klasių savybės, klasė, ryšiai tarp klasių. Ontologija yra sudaryta iš klasių, ryšių tarp klasių ir klasių savybių. Klasė gali turėti daug savybių ir daug ryšių su kitomis klasėmis. Žodžiai sudaro frazę, frazė – sakinį.



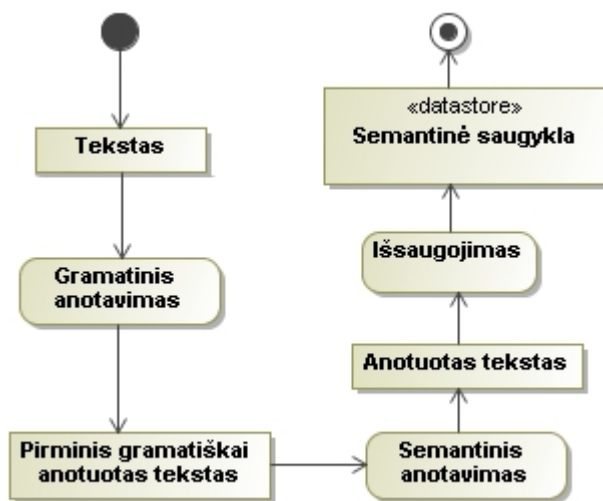
1.1 pav. Sporto įvykių dalykinės srities esybių modelis

Semantinis anotavimas tai – teksto pažymėjimas ontologijos konceptu ir įtraukimas į ontologiją [7]. 1.2. paveiksle vaizduojama: teksto fragmentas, ontologijos, esančios semantinėje saugykloje, klasės ir jų tarpusavio ryšiai.



1.2 pav. Semantinis anotavimas [7]

Semantinio anotavimo procesą galima suskirstyti į tris etapus: gramatinis nagrinėjimas, teksto žymėjimas, išsaugojimas (1.3 paveikslas).



1.3 pav. Bendras semantinio anotavimo procesas

Pirmame etape tekstas išnagrinėjamas gramatiškai, taip išskiriant tekste esančias kalbos dalis ir gaunant tekste esančių sakinių struktūrą. Gramatinį nagrinėjimą galima suskirstyti į keletą fazių:

- Teksto segmentavimas (teksto struktūros anotavimas);
- Įvardintų esybių atpažinimas;
- Morfologinis anotavimas (POS);
- Sintaksinis anotavimas [29];

Tekstas susegmentuojamas į pastraipas, sakinius ir žodžius.

Prieš anotuojant morfologiškai, atpažįstamos įvardintos esybės. Morfologinio anotavimo metu kiekvienam teksto žodžiui suteikiama gramatinė pažyma, t.y. nurodomos gramatinės žodžio savybės: kokia tai kalbos dalis, giminė, linksnis ir panašiai. Sintaksinio anotavimo metu atpažįstami ir pažymimos sakinio dalys, suformuojamas sintaksinis medis.

Semantinio anotavimo tikslumas priklauso nuo to, kaip gerai gramatiškai išnagrinėtas tekstas. Jeigu tekstas bus išnagrinėtas klaidingai, tai atsilies ir semantinio anotavimo tikslumui.

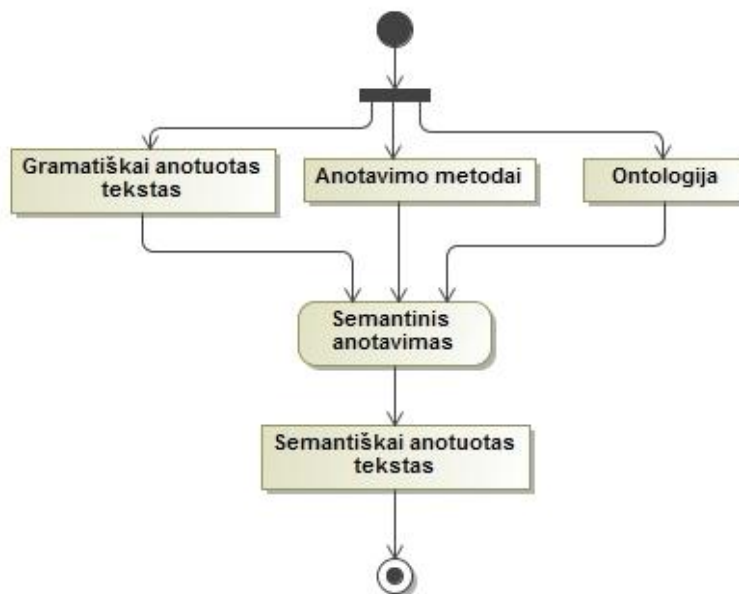
Antrame etape gramatiškai išnagrinėtas tekstas žymimas ontologijos conceptais.

Trečiame etape ontologijos conceptu pažymėti žodžiai įtraukiami į ontologiją, o anotuotas tekstas išsaugomas.

1.4 paveiksle pavaizduotos semantinio anotavimo proceso įeiga ir išeiga. Anotavimo procesui reikalingas morfologiškai ir sintaksiškai anotuotas tekstas, anotavimo metodai, kuriais vadovaujantis atliekamas semantinis anotavimas, ir ontologija, pagal kurią anotuojamas tekstas.

Semantinis anotavimas gali būti vykdomas trimis būdais: rankiniu, pusiau automatiniu ir automatiniu. Rankinis anotavimas – esamos sintaksės išteklių transformavimas į tarpusavyje

susijusias žinių struktūras su atitinkama informacija. Pusiau automatinis anotavimas – kai anotavimo sistemos veikia su tam tikru žmogaus įsikišimu, kuriame nors anotavimo etape. Automatinis anotavimas – visiškai automatizuotas anotavimas.



1.4 pav. Bendra semantinio anotavimo proceso įeiga ir išeiga

1.4. Vartotojų analizė

1.4.1. Vartotojų aibė, tipai ir savybės

Semantinio anotavimo procesas liečia dviejų tipų vartotojus: tiesioginius ir netiesioginius. Tiesioginiai vartotojai tiesiogiai susiduria su semantinio anotavimo procesu, jame dalyvauja. Netiesioginiai – nedalyvauja pačiame procese, tačiau pajaučia jo teikiamą naudą. Potencialių vartotojų kvalifikacijos pagal tipus išvardintos žemiau esančioje 1.1 lentelėje.

1.1 lentelė. Vartotojai

Tipas	Kvalifikacija	Veikla
Tiesioginiai	IT specialistas	Kuria naujas, lietuvių kalbos analize pagrįstas sistemas ar tam tikrų sistemų funkcijas.
	Kompiuterinės lingvistikos specialistas	Perkelia lietuvių kalbą į skaitmeninę erdvę, kaupia ir kuria kalbos išteklius ir įrankius.
	Lietuvių k. specialistas	Įvairiais aspektais tiria lietuvių kalbą.
Netiesioginiai	Paprastas interneto vartotojas	Ieško informacijos internetinėje erdvėje.

1.4.2. Vartotojų tikslai ir problemos

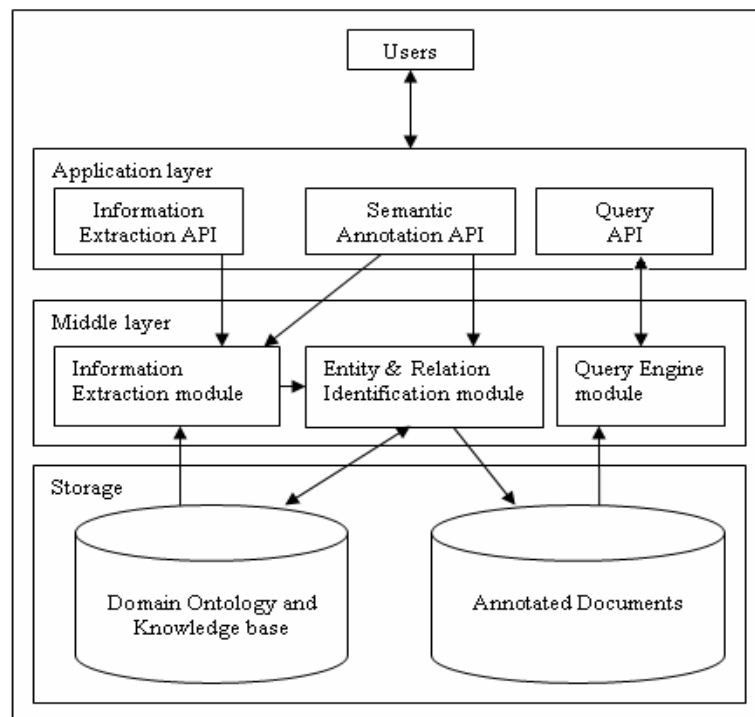
Siekiant geriau suprasti vartotojų poreikius, sudaryta lentelė „Vartotojų tikslai ir problemos“, kuri pateikiama žemiau esančioje 1.2 lentelėje.

1.2 lentelė. Vartotojų tikslai ir problemos

Vartotojas	Problema	Kaip viskas vyksta dabar	Tikslas
Lietuvių k. specialistas	Nėra sudaryto semantiškai anotuoto lietuviško tekstyno.	Yra iš dalies morfologiškai anotuotas tekstynas.	Sudaryti sąlygas efektyviau tikri lietuvių kalbos semantiką.
IT specialistas	Paieškos sistemose naudojami ne naujausi sprendimai.	Paieška vykdoma naudojant tikimybių metodą, kai rezultatai atrenkami ne pagal prasmę o pagal tikimybę.	Sukurti efektyvesnes paieškos sistemas.
Paprastas interneto vartotojas	Neefektyvios esamos paieškos sistemos.	Paieškos rezultatai ne visada atitinka lūkesčius.	Naudotis efektyvesnėmis paieškos sistemomis.
Kompiuterinės lingvistikos specialistas	Rankiniu būdu anotuojama tik maža dalis teksto.	Specialistai, skaitydami tekstą, patys jį anotuoja.	Automatizuoti teksto, parašyto lietuvių k., semantinio anotavimo procesą.

1.5. Esamų sprendimų, skirtų semantiniam anotavimui, analizė

Žemiau (1.5 pav.) pateikiama bendrosios semantinio anotavimo platformų architektūros schema. Iš jos matome, kad anotavimo platformos būna sudarytos iš trijų sluoksnių. Taikomojo lygmens sluoksnis (*angl. Application layer*) yra atsakingas už galutinio vartotojo sąsają ir teikiamas paslaugas. Jis yra pagrindinis taikomųjų programų sąsajos sluoksnis. Viduriniame sluoksnyje (*angl. Middle layer*) yra pagrindiniai komponentai atliekantys pagrindinius veiksmus. Saugojimo sluoksnis (*angl. Storage*) skirtas ilgalaikių duomenų pvz. ontologijų, dokumentų anotacijų ir žinių bazių (*angl. Knowledge base*) saugojimui ir jo valdymui.

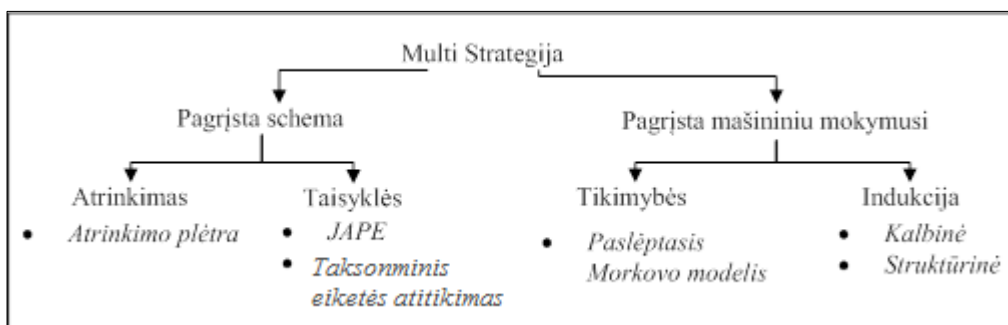


1.5 pav. Bendroji semantinio anotavimo platformų architektūra [14]

Semantinės anotacijos platformos gali būti klasifikuojamos pagal naudojamus anotacijos metodus. Yra dvi pagrindinės semantinių anotavimo platformų kategorijos – pagrįsta schema ir pagrįsta mašininiais mokymais. Platformos, kurių veikimas pagrįstas schema, naudoja dviejų tipų metodus: atrinkimo ir taisyklių. Dauguma atrinkimo metodų veikia Brin metodo principu, kai yra žinomas pradinis apibrėžtas subjektų rinkinys ir tekstas yra nuskaitomas ieškant modelių, kuriems tie subjektai priklauso [12]. Taisyklėmis paremtas metodas naudoja rankiniu būdu parašytas taisykles, skirtas surasti subjektus tekste. Taisyklėmis paremtas anotavimas gali būti tikslesnis nei paremtas mašininiais mokymais, tačiau reikalauja didelio dėmesio aprašant taisykles, ir labai gero srities išmanymo.

Metodas pagrįstas mašininiais mokymais yra skirstomas į paremtą tikimybėmis ir indukcija. Semantinės anotacijos platformos, pagrįstos tikimybėmis, pagal tam tikrus metodus apskaičiuoja kokia tikimybė, kad pvz. tam tikra teksto dalis atitiks tam tikrą ontologijos konceptą. Platforma, pagrįsta indukcija, pritaiko jau žinomas subjektų savybes kitiems subjektams.

Prireikus galima naudoti abiejų kategorijų metodus, taip išnaudojant kiekvieno stiprybes.[12] Platformų klasifikacija pateikiama žemiau esančiame 1.6 paveiksle, o platformos palyginimas – 1.3 lentelėje.

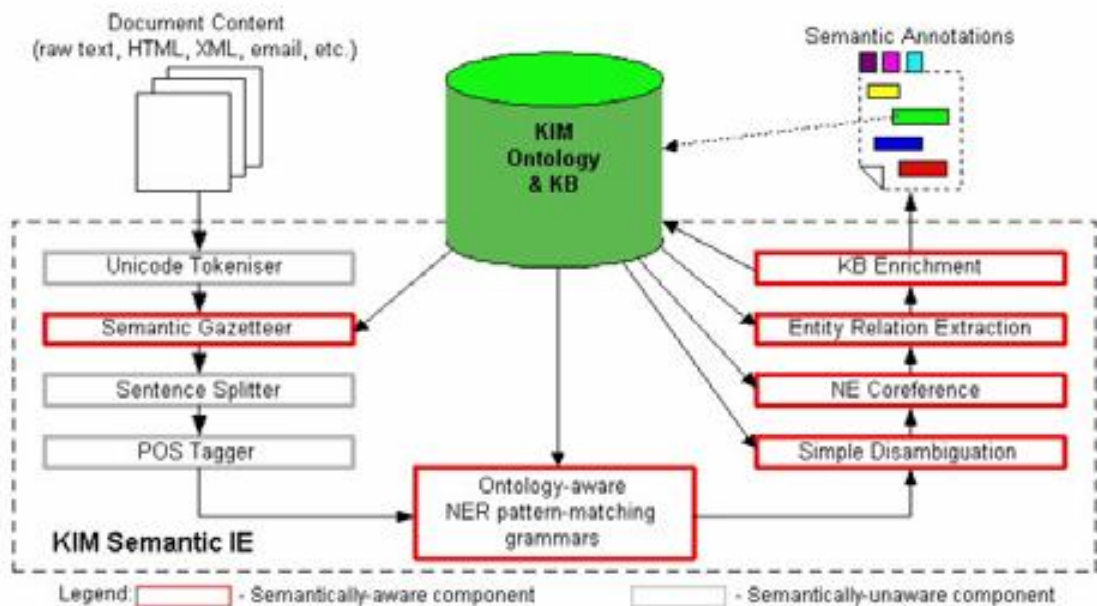


1.6 pav. Semantinio anotavimo platformų klasifikacija [12]

1.5.1. KIM

KIM (*Knowledge and Information Management*) yra semantinio anotavimo platforma, teikianti žinių ir informacijos valdymo infrastruktūrą ir paslaugas skirtas semantiniam anotavimui, indeksavimui bei nestruktūrizuotam ar pusiau struktūrizuotam turiniui [8]. Joje naudojamas anotacijos metodas pagrįstas taisyklėmis.

Žemiau esančiame 1.7 paveiksle pavaizduota, kaip vyksta semantinis anotavimas KIM platformoje.



1.7 pav. Semantinis anotavimas naudojantis KMI [13]

1.5.2. MnM

MnM – anotavimo įrankis, suteikiantis galimybę automatiškai ar pusiau automatiškai anotuoti semantinio turinio internetinius puslapius. Jis integruoja interneto naršyklę su ontologijų redaktoriumi.[9]

Pagrindinės penkios veiklos palaikomos MnM:

- *Naršymas*. Konkrečių žinių komponentai yra parenkami iš žinių modelių bibliotekos ontologijos serveryje.
- *Žymėjimas*. Pasirinkta grupė žinomų komponentų yra parenkama suformuoti naujam IE mechanizmo pagrindui. Dokumentų tekstynas yra pažymimas rankiniu būdu.
- *Mokymasis*. Mokymosi algoritmas yra paleidžiamas per pažymėtą tekstyną išmokti ištraukimo taisykles.
- *Bandymas*. IE mechanizmas yra paleidžiamas per bandomąjį tekstyną įvertinti jo tikslumui ir prisiminti priemonėms.
- *Ištrauka*. IE mechanizmas yra pasirenkamas ir paleidžiamas per dokumentų rinkinį ir išgaunama informacija, kuri vėliau siunčiama į ontologijos serverį, kur yra vartotojo pasirinkta ontologija [14].

1.5.3. GATE

GATE yra įrankis skirtas:

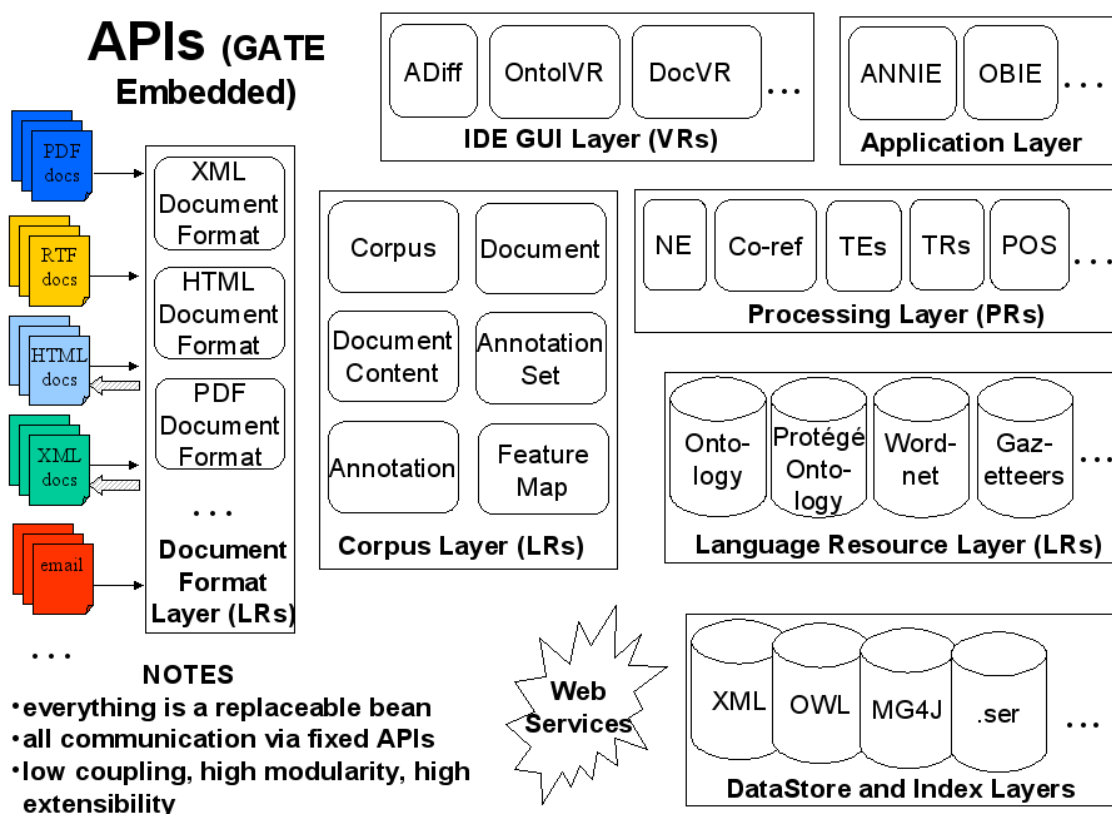
- mokslininkams atlikti eksperimentus, susijusius su žmogaus kalba;
- bendrovėms, kuriančioms programas su kalbos apdorojimo komponentėmis;
- kalbos skaičiavimo kursų dėstytojams ir studentams.

Ši sistema buvo naudojama daugelyje kalbos apdorojimo projektų [10].

GATE yra:

- GATE Developer4: integruota kūrimo aplinka kalbos apdorojimo komponentų rinkinys su labai plačiai naudojama informacijos išgavimo sistema ir plačiu kitų įskiepių rinkiniu;
- GATE Teamware: bendra anotacijos aplinka, skirta įmonių semantinės anotacijos projektams;
- GATE Embedded: objektų biblioteka optimizuota įtraukimui į įvairias programas, turinčias priėjimą prie visų paslaugų, naudojant GATE Developer ;
- Aukšto lygio organizacinis kalbos apdorojimo programinės įrangos sudėties vaizdas;
- Stabilus ir patogus aptarnavimo paslaugų kūrimo procesas.

Viena iš priežasčių, kodėl GATE yra sėkmingai naudojama platforma, yra tai, kad ji suskirstyta į daugkartinio naudojimo dalis (naudojant Java komponentų modelį). Žemiau esančiame 1.8 paveiksle pavaizduotas GATE sudarančių komponentų modelis.



1.8 pav. Komponentų modelis [16].

1.5.4. OntoMat Annotizer

OntoMat Annotizer – interaktyvus tinklalapių anotacijos įrankis. Jis leidžia anotatoriui, naudojant „drag and drop“ sąveiką, paryškinti atitinkamas tinklalapio dalis ir sukurti naujus atvejus.

Jis palaiko metaduomenų kūrimą gyvenimo ciklo etape. Planuojama, kad ateityje jame bus pateikti informacijos išgavimo įskiepai, parodantys, kurios teksto dalys yra svarbios anotavimui [11].

1.5.5. Semantinio anotavimo platformų palyginimas

Žemiau esančioje 1.3 lentelėje pateikiamas jau esamų semantinio anotavimo platformų palyginimas.

1.3 lentelė. Semantinio anotavimo platformų palyginimas

Platforma	Metodas	Naudojamos ontologijos	Pritaikymas lietuvių kalbai	Atnaujinimas	Kaina	Prieinamumas
<i>KMI</i>	Taisyklės	PROTON ir vartotojo	Negalimas	2009	Nemokama naudojant ne komerciniais tikslais.	Ne
<i>MnM</i>	Modelio atrinkimas	KMi	Nedidelis	2004	Nemokamas	Atviro kodo
<i>GATE</i>	Taisyklės	Vartotojo	Galimas	2011	Nemokamas	Atviro kodo
<i>OntoMat Annotizer</i>	Modelio atrinkimas	Vartotojo	Negalimas	2004	Nemokamas	Ne

Kiekvienos kalbos struktūra yra unikali, todėl kiekvienai kalbai reikia jai pritaikyto semantinio anotavimo modelio ar taisyklių. Platformos KMI ir Onto Mat Annotizer nėra atviro kodo, todėl jas pritaikyti lietuvių kalbai būtų sudėtinga. MnM projekto vystymasis yra sustojęs. Bandyti jį pritaikyti lietuvių kalbai nėra tikslinga. GATE projektas yra nuolat plečiamas, atnaujinamas, turintis išsamią dokumentaciją ir prieinamas kiekvienam interneto vartotojui. Jis susideda iš komponentų, kuriuos galima panaudoti, pritaikyti kitoms kalboms.

1.6. Esamų ontologijų kūrimo įrankių analizė

Ontologija, semantinio anotavimo procese, užima svarbią vietą. Be jos nebūtų galima atlikti semantinio anotavimo. Vienas iš darbo uždavinių yra sukurti sporto įvykių ontologiją. Ontologijos kūrimui reikalingi tam pritaikyti įrankiai. Yra sukurta nemažai įrankių, skirtų kurti ontologijoms. Žemiau esančiuose skyriuose pateikta populiariausių įrankių analizė.

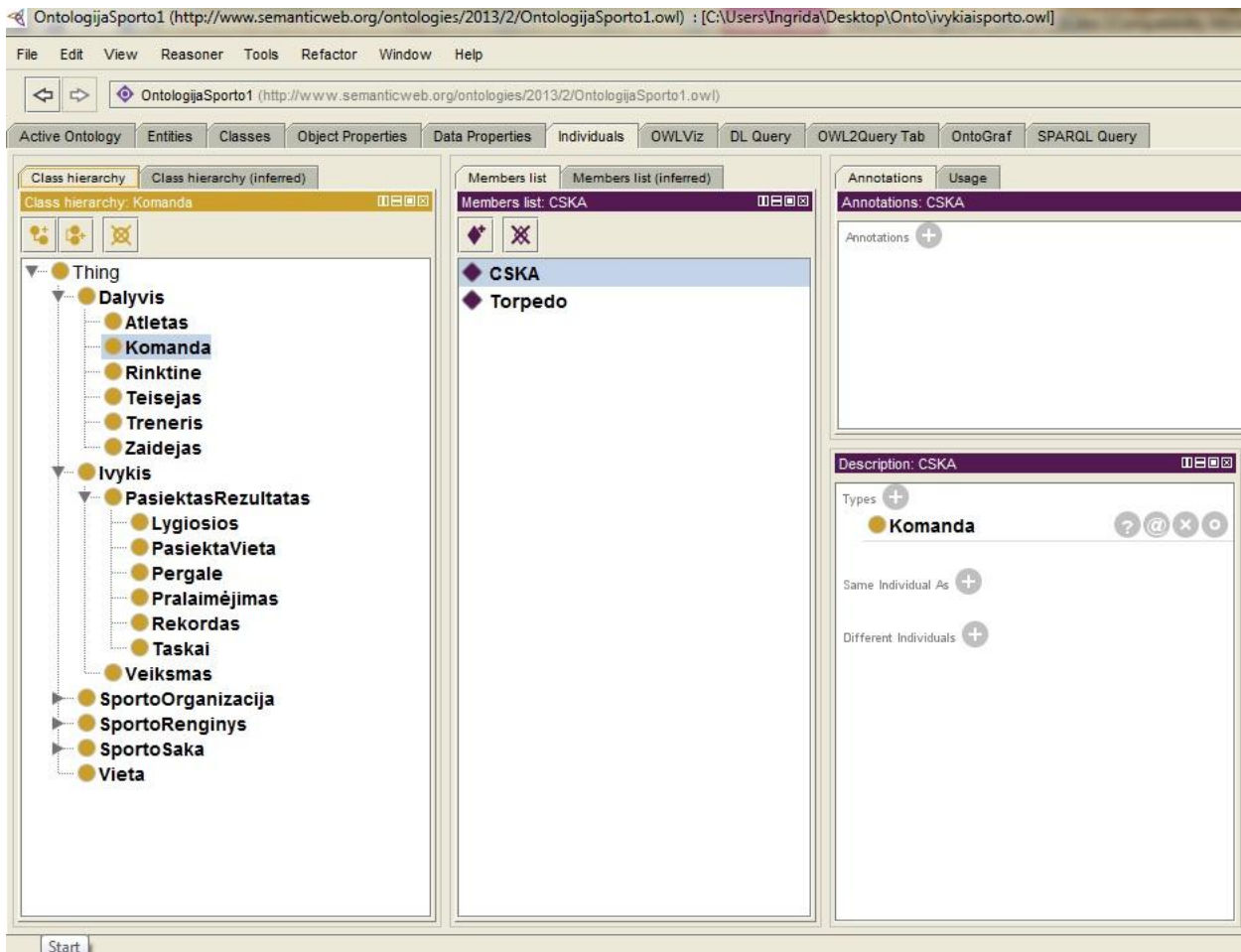
1.6.1. Protégé

Protégé – yra nemokama, atviro kodo platforma, kuri, augančiai vartotojų bendruomenei, suteikia rinkinį įrankių, skirtų kurti srities modelius, ir žiniomis paremtas taikomąsias programas su ontologijomis. Tai turtingas rinkinys žiniomis paremto modeliavimo struktūrų ir veiksmų, kurie padeda kurti, peržiūrėti ir manipuluoti įvairių formatų ontologijomis. Protégé gali būti pritaikytas suteikiant sričiai draugišką palaikymą, kuriant žinių modelius ir įvedant duomenis. Įrankio

funktionalumas gali būti padidintas naudojant įskiepius ir java pagrįstomis aplikacijų programavimo sąsajomis (API). Įrankio Protégé langas pateikiamas žemiau esančiame 1.9 paveiksle.

Protégé palaiko du pagrindinius ontologijų modeliavimo būdus:

- Protégé-Frames redaktorius leidžia vartotojams kurti ir užpildyti ontologijos karkaso pagrindu, laikantis Open Knowledge Base Connectivity protokolo (OKBC).
- Protégé-OWL redaktorius leidžia vartotojams kurti ontologijos semantiniam žiniatinkliui, visų pirma OWL (W3C's Web Ontology Language) kalba. OWL ontologija gali būti aprašoma klasės, savybės ir jų egzemplioriai. [20]



1.9 pav. Protégé langas

1.6.2. TopBraid Composer

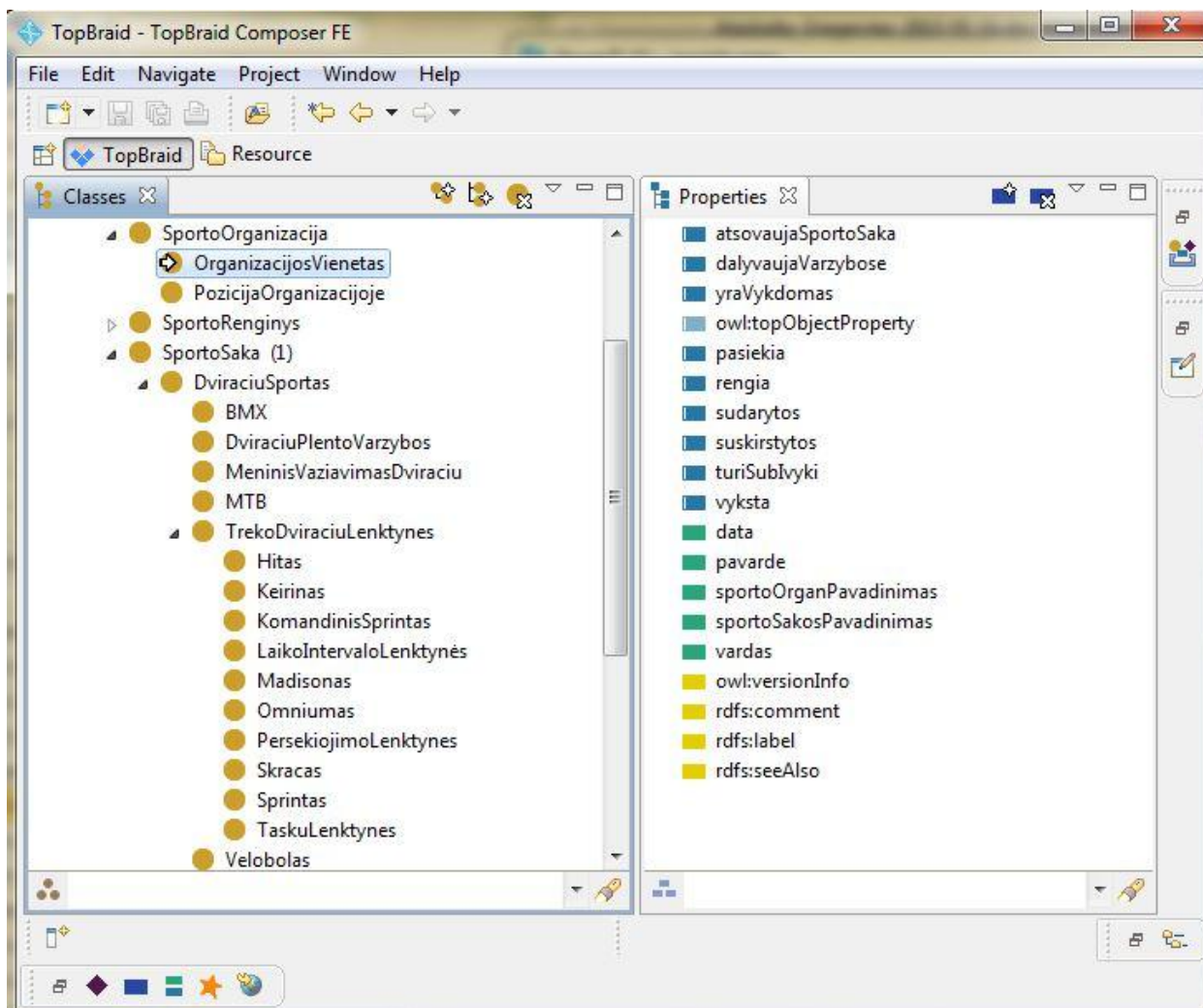
TopBraid Composer kūrėjai sako, kad jų įrankis yra galingiausia pasaulyje modeliavimo ir integruota kūrimo aplinka, visiškai atitinkanti W3C standartus. Composer siūlo pilną paramą, kuriant, valdant ir testuojant ontologijas ir susijusius duomenis. Kaip dalis TopBraid rinkinio, Composer suteikia išsamią integruotą kūrimo aplinką (IDE) plėtoti semantinius sprendimus, kurie gali būti integruoti su skirtingomis programomis ir duomenų šaltiniais.

Yra pateikiamos trys įrankio versijos:

- Maestro Edition – tai integruota kūrimo aplinka, pilnai pritaikyta programuoti žiniatinklio taikomąsias programas pagrįstas TopBraid Live platforma;

- Free Editions – yra įrankis, skirtas pradedantiesiems kurti RDF/OWL failus bei leisti SPARQL užklausas;
- Standard Editions – yra funkcionalesnė nei Free Editions. Ji turi vizualius redaktorių RDF grafams ir klasių diagramoms. [21]

TopBraid Composer Free Editions langas pateikiamas žemiau esančiame 1.10 paveiksle.



1.10 pav. TopBraid Composer Free Editions langas

1.6.3. Ontologijų kūrimo įrankių palyginimas

Žemiau esančioje 1.4 lentelėje pateikiamas ontologijų kūrimo įrankių palyginimas.

1.4 lentelė. Ontologijų kūrimo įrankių palyginimas

Platforma	Kalbų palaikymas	Antologijos grafinis atvaizdavimas	SPARQL užklausų rašymas	Kaina	Prieinamumas
Protégé	OWL, RDF, Turtle.	Yra, galima peržiūrėti ir grafo pavidalu.	Yra	Nemokama naudojant ne komerciniais tikslais.	Atviro kodo
TopBraid	OWL, RDF, N-Triple,	Yra.	Yra	Yra nemokama	Ne

Composer	Turtle			bandomoji versija.	
----------	--------	--	--	--------------------	--

Ontologijos kūrimui pasirinktas Protégé įrankis, nes jis yra nemokamas ir prieinamas kiekvienam, taip pat nuolat atnaujinamas ir tobulinamas, palaiko pagrindines ontologijų kūrimo kalbas ir turi patogią grafinio atvaizdavimo, kai ontologija atvaizduojama grafo pavidalu, funkciją.

1.7. Esamų sporto ir įvykių sričių ontologijų analizė

Siekiant geriau įsigilinti į ontologijas ir jų sandarą atliekama jau egzistuojančių ontologijų analizė. Ontologija – dalykinės srities terminų ir santykių tarp jų, specifikacija. Ontologijos yra sudarytos iš klasių, klases aprašo savybės[4]. Tyrimo objektu buvo pasirinktas sporto įvykių atpažinimas lietuviškame tekste ir tokių tekstų semantinis anotavimas. Todėl geriausia būtų analizuoti ontologijas, susijusias su sporto įvykiais. Kadangi sporto įvykių ontologija dar nėra sukurta, todėl buvo nuspręsta analizuoti dvi ontologijas: sporto ir įvykių. Pasirinkta BBC Sporto ontologija yra realiai naudojama BBC puslapyje skelbti sporto įvykiams [24]. Analizuojama Įvykių ontologija, apibrėžianti pagrindinę įvykio koncepciją. Ši ontologija yra išsamiai aprašyta, ją galima parsisiųsti ir naudoti.

Žemiau esančiuose poskyriuose pateiktos išnagrinėtos pasirinktos ontologijos.

1.7.1. BBC Sporto ontologija

Ši sporto ontologija yra paprasta ontologija, skirta skelbti duomenis apie konkurencingus sporto renginius. Šioje ontologijoje naudojami terminai leidžia paskelbti duomenis apie:

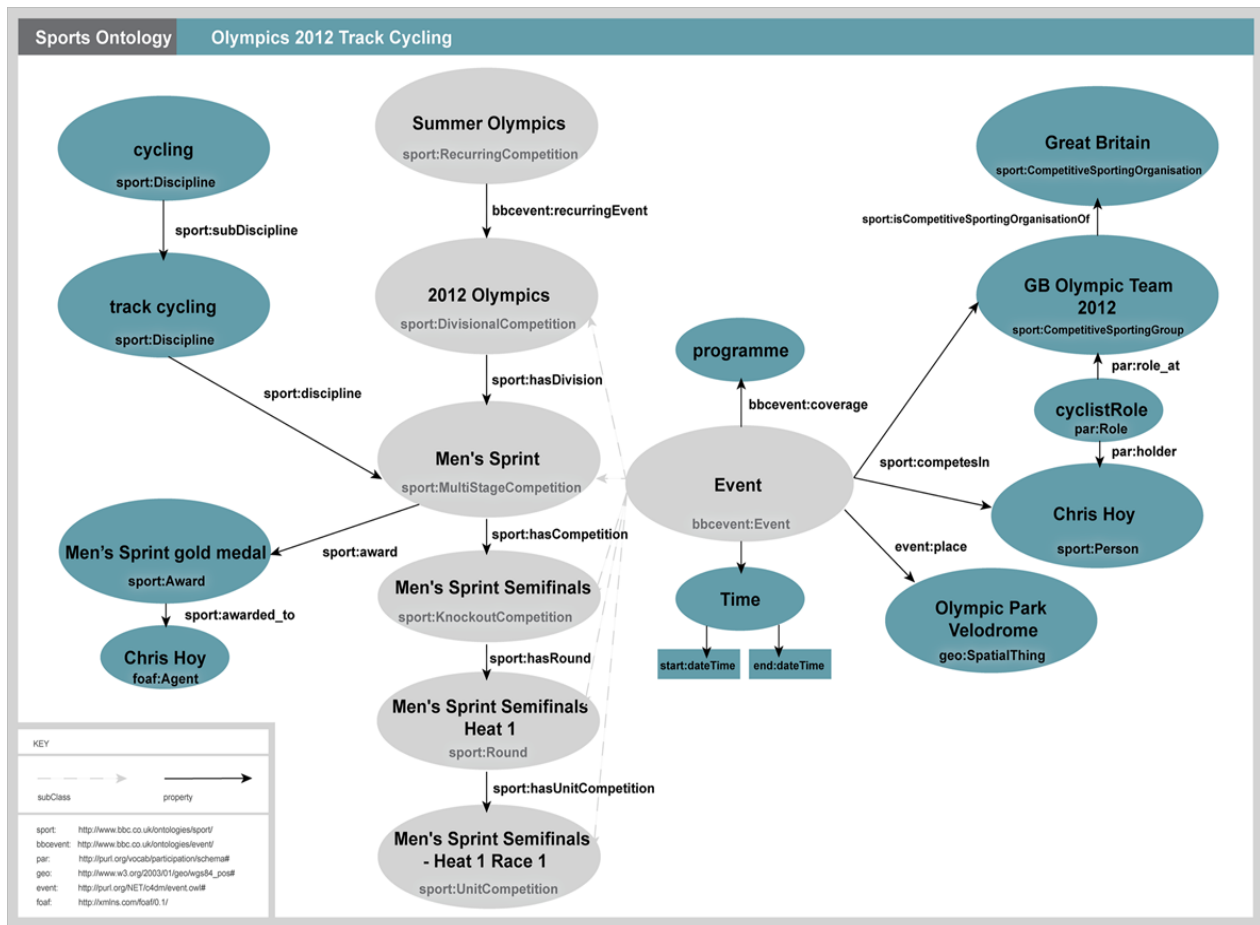
- Struktūrą sporto turnyrų, kaip renginių seką;
- Rungtynėse konkuruojančius dalyvius;
- Su kokia disciplina susijęs įvykis;
- Apdovanojimas susijęs su konkurencingu sporto renginiu ir kaip jis gautas ir t.t..

Nors ji yra dėl konkrečių BBC naudojimo atvejų, tačiau Sporto ontologija turėtų būti naudojama skelbti įvairių konkurencingų sporto renginių duomenims. Buvo stengtasi pabandyti užtikrinti suderinamumą su pagrindinių ontologijų naudojimu. Visų pirma, jame panaudota daug sprendimų iš Įvykių ontologijos.

Sporto ontologijos kūrimo metu buvo peržiūrėti jau esami RDF duomenų modeliavimo ir publikavimo sprendimai. Apibendrintai tai:

- Įvykių ontologija;
- SportsML;
- Vietovardžių ontologija.

Žemiau esančiame 1.11 paveiksle pateikta schema, vaizduojanti ryšius tarp pagrindinių ontologijos klasių, pritaikius ją Olimpiados dviračių rungtynėms.



1.11 pav. Olimpiados 2012 dviračių rungtynių modelis [18]

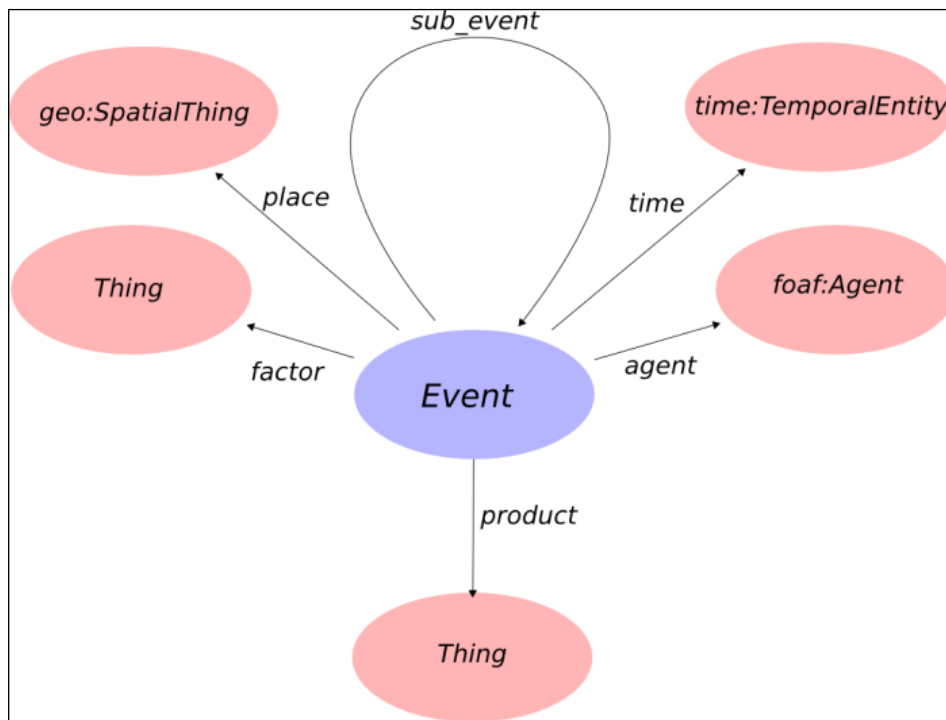
Sporto ontologijos klasės:

- *Competition* – konkurencingas sporto renginys;
- *Competition Type* – sporto varžybų tipas. Pvz. vidaus lygio ar Europos taurės rungtynės;
- *Competitive Sporting group* – grupė agentų (dalyvių) susijusių su sporto varžybomis. Dvi grupės su tais pačiais nariais nebūtinai turi būti ta pati grupė. Pvz.: 2010 Pasaulio Taurės Anglijos komanda ir 2012 Europos čempionato komanda;
- *Competitive sporting organisation* – sporto organizacija, kuri dalyvauja konkurencinguose sporto renginiuose. Pvz.: Manchester United arba Didžiosios Britanijos olimpinė komanda;
- *Divisional competition* – tai sporto renginys, kuris yra padalintas į keletą dalių, kurių kiekviena yra apdovanojama.
- *Event gender* – kokios lyties dalyviai dalyvauja sporto įvykyje;
- *Football manager role* – futbolo komandos kapitono vaidmuo;

- *Football player role* – futbolo žaidėjo vaidmuo;
- *Group competition* – grupės varžybos yra tokios varžybos, kuriose į sekantį etapą išrenkamos geriausiai grupėse pasirodžiusios komandos;
- *Knock out competition* – eliminacinės varžybos;
- *League competition* – lygos varžybos;
- *Match* – Rungtynės. Varžybų vienetų tipas. Pvz.: futbolo ar teniso varžybos;
- *Multiround competition* – varžybos, kurios yra organizuojamos kaip raundų seka;
- *Multistage competition* – kelių etapų varžybos yra tokios, kurios organizuojamos kaip kelių etapų seka. Pvz.: Pasaulio futbolo taurės varžybos;
- *Recurring competition* – pasikartojančios sporto varžybos;
- *Round* – tai viena ar daugiau varžybų, kurios yra dalis Multiround varžybų;
- *Session* – sesija yra pavadintas varžybų intervalas;
- *Sport governing body* – sporto valdymo organas yra sporto organizacija, kuri atlieka įvairių sankcijų teikimo ir reguliavimo funkcijas. Pvz. FIFA;
- *Sporting organisation* – sporto organizacija yra bet kokia organizacija, kuri dalyvauja įvairiuose sporto renginiuose;
- *Sports discipline* – sporto disciplina;
- *Unit competition* – yra žemiausio lygio varžybos.

1.7.2. Bendroji įvykių ontologija

Yves Raimond ir Samer Abdallah sukurtą ontologiją apibrėžia vieną pagrindinę įvykio koncepciją. Įvykis gali turėti vietą, laiką, aktyvų agentą, veiksnius ir rezultatus. Žemiau esančiame 1.12 paveiksle pateiktas Įvykių ontologijos modelis.



1.12 pav. Įvykių ontologijos modelis [19]

Įvykio ontologijos klasės:

- *Event* (Įvykis) – savavališka klasifikacija laiko/erdvės regione pagal agento (veikėjo) suvokimą;
- *Factor* (Veiksny) – viskas, kas naudojama įvykyje kaip veiksnys;
- *Product* (Rezultatas) – viskas, ką sukuria įvykis.

Ontologijos savybės:

- *agent* (agentas, veikėjas) – aktyvus agentas, susijęs su įvykiu (asmuo, kompiuteris ir pan.);
- *factor* (veiksny) – susiję su įvykiu pasyvūs veiksniai (įvykis, instrumentas, abstrakti priežastis);
- *literal_factor* – susijęs su įvykiu veiksnys, kurį galima apibūdinti kaip pažodinį. Ši savybė neturėtų būti naudojama viena, tačiau turėtų būti sujungta su kita, labiau specifine savybe;
- *place* (vieta) – susieja įvykį su erdviniu objektu;
- *product* (rezultatas) – susieja įvykį su kažkuo, kad sukurta ar išgauta jo metu (garsas, pyragas ir pan.);
- *sub_event* (sub įvykis) – ši savybė leidžia padalinti sudėtingą įvykį į keletą paprastesnių;
- *time* (laikas) – susieja įvykį su laiko objektu.

1.7.3. Esamų ontologijų panaudojimas

BBC sporto įvykių ontologija labiausiai buvo pritaikyta 2010 m. Pasaulio Taurės čempionatui. Todėl naudojantis ja gali būti sudėtinga anoutuoti tekstus su įvykiais iš sporto šakų, turinčių specifines sąlygas: pvz. skirtingos rungtys penkiakovėj ir kaupiami taškai. Taip pat gali kilti problemų su kitais čempionatais, varžybomis ir t.t., kurių struktūra skiriasi nuo Pasaulio Taurės čempionato. Nors apie BBC ontologiją yra nemažai rašoma, ji nuolat atnaujinama, tačiau pačios jos parsisiųsti nėra galima, priešingai nei nagrinėtos Įvykių ontologijos. Įvykių ontologija, netinkama fiksuoti sporto įvykiams, nes yra per daug abstrakti, dėl specifikos trūkumo nebūtų galima atpažinti konkrečių įvykių. Todėl būtų nepatogu ją naudoti ir sudėtinga gauti tikslius rezultatus, atliekant paiešką tarp sporto įvykių.

1.8. Ontologijos kūrimo metodika

Ontologijos kūrimas – sudėtingas ir gerų dalykinės srities žinių reikalaujantis procesas. Todėl geriausia ontologiją kurti vadovaujantis jau sukurta metodika. Kadangi nėra vienos universalios ir geriausios metodikos ontologijoms kurti [17, 41, 48] , šiame darbe kursiu ontologiją, vadovaudamasi Stanford University dėstytojų aprašyta instrukcija, kaip sukurti ontologiją.

Pagrindiniai žingsniai:

1. *Nustatoma kuriamos ontologijos apimtis ir sritis.* Žemiau pateikti pagrindiniai klausimai, kurie turi būti atsakyti prieš kuriant ontologiją, ir atsakymai apie mano ontologiją:
 - Kokią sritį ontologija apims? – Ontologija bus skirta sporto įvykiams.
 - Kam bus naudojama ontologija? – Ji bus naudojama paieškai atlikti.
 - Į kokius klausimus ontologija turės pateikti atsakymus?: Ontologija turės pateikti atsakymus į klausimus apie įvairių sporto šakų įvykius, juose dalyvaujančius asmenis, jų veiksmo vietą ir pan.
 - Kas naudosis ir prižiūrės ontologiją? – Sistemoje paiešką galės atlikti paprasti vartotojai. Ją panaudoji, anoutuojant tekstą, galės tos srities specialistai.
2. *Apsvarstoma galimybė ar galima pasinaudoti jau esamomis ontologijomis.* Visada verta pasidomėti, ar kas nors padarė ką nors panašaus dominančioje srityje, ką būtų galima patobulinti ir pritaikyti patiems, iškilusioms problemoms spręsti. Šiuo atveju galima pasinaudoti jau egzistuojančiomis BBC Sporto ir Įvykių ontologijomis.
3. *Išvardijami svarbiausi ontologijos terminai.* Sudaromas svarbiausių terminų sąrašas. Kitu du žingsniai – klasių sudarymas ir jų savybių aprašymas yra labai susiję. Sunku padaryti vieną iš jų ir tada daryti kitą. Todėl, įvedus kelias klases, apibrėžiamos jų savybės, tada įvedamos kitos klasės ir t.t..
4. *Apibrėžiamos klasės ir jų hierarchija.*

5. *Apibrėžiamos klasių savybės – slotai.*
6. *Nustatomos klasių ypatybės ir apibrėžimai: kardinalumai, tipai [17].*

1.9. Darbo tikslas ir uždaviniai

Tikslas – sudaryti sąlygas semantiškai anotuoti su sporto įvykiais susijusius lietuviškus tekstus ir tokiu būdu prisidėti prie lietuvių kalbos įsitvirtinimo pasauliniame semantiniame tinkle.

Uždaviniai:

1. Išanalizuoti semantinio anotavimo procesą;
2. Išanalizuoti teksto anotavimo įrankius, formatus ir standartus;
3. Sukurti sporto įvykių ontologiją;
4. Specifikuoti sporto įvykių atpažinimo tekste taisykles;
5. Parengti metodiką, skirtą atpažinti sporto įvykius lietuviškame tekste;
6. Ištirti sukurtą sporto įvykių ontologiją ir metodiką.

1.10. Siekiamo sprendimo apibrėžimas

Magistrinio darbo metu bus siekiama sukurti įvykio ontologiją, kurią būtų galima naudoti lietuviškų tekstų semantiniam anotavimui. Taip pat siekiama sukurti metodiką, kuria vadovaujantis būtų galima sukurti priemones, reikalingas automatiškai atpažinti sporto įvykius lietuviškame tekste. Metodika kuriama atsižvelgiant į lietuvių kalbos taisykles, struktūrą.

Kuriant sporto įvykių ontologiją bus vadovaujama pagrindiniais įvykių ontologijų principais:

1. Įvykius galima ir reikia skaidyti į smulkesniu įvykius;
2. Įvykiai turi turėti:
 - subjektą, kuris inicijuoja įvykį;
 - objektą, kurį paveikia įvykis;
 - rezultatą.
3. Įvykiai gali turėti:
 - vietą, kurioje vyksta įvykis;
 - datą, laiką arba laiko tarpą, kuriuo vyko įvykis.

Pagrindinis sukurto sprendimo kokybės kriterijus – ar, sukurtoje ontologijoje atliekama paieška pagal ontologijos kūrimo metu apsibrėžtus klausimus ir, ar vadovaujantis aprašytomis taisyklėmis, galima aptikti sporto įvykį tekste.

1.11. Analizės išvados

1. Vartotojų analizės metu išsiaiškinti proceso vartotojai (IT specialistas, kompiuterinės lingvistikos specialistas, lietuvių k. specialistas, paprastas interneto vartotojas) ir šiems vartotojams kylančios problemos.

2. Problemos kylančios dėl automatinio semantinio anotavimo lietuvių kalba nebuvimo: ne visada reikiamus rezultatus pateikiančios esamos paieškos sistemos, rankiniu būdu anotuojama tik maža dalis teksto, nėra semantiškai anotuoto tekstyno, kuris leistų geriau analizuoti lietuvių kalbos semantiką;
3. Išanalizuoti keturi panašūs sprendimai: KMI, MnM, GATE, OntoMat Annotizer. Iš kurių lietuvių kalbai, dėl slavų ir lietuvių kalbų panašumo, komponentų naudojimo, labiausiai pritaikomi yra MnM ir GATE;
4. Nuspręsta gilintis į sporto įvykių laimėjimų/pralaimėjimų atpažinimą ir sukurti atitinkamą metodiką. Vadovaujantis šia metodika būtų galima atpažinti sporto įvykius lietuviškam tekste;
5. Išanalizuotos BBC Sporto ir Įvykių ontologijos ir nuspręsta kurti sporto įvykių ontologiją, remiantis jau sukurtomis BBC sporto ir Įvykių ontologijomis. Nes įvykių ontologija pateikia bendriausią įvykio koncepciją, o BBC sporto ontologiją – sporto renginių struktūrą.

2. SPRENDIMO REIKALAVIMŲ SPECIFIKACIJA

2.1. Reikalavimų metodikai specifikacija

Kuriame metodika bus naudojama sporto įvykių anotavimui, todėl ji turi turėti:

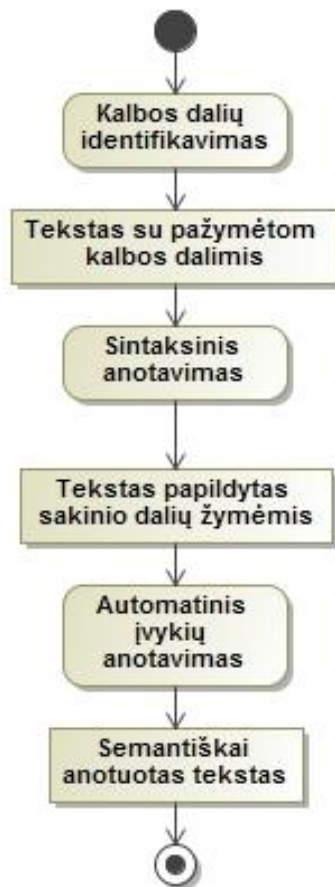
- Formaliai ir natūralia kalba aprašytas taisykles, skirtas lietuviškame tekste atpažinti sporto įvykius;
- Tinkamai sudarytą sporto įvykių ontologiją;
- Aprašytą taisyklių taikymo tvarką, pagal kurią, naudojantis sukurta ontologija, galima įvykdyti metodiką.

Automatinis sporto įvykių atpažinimo turi vykti tokia tvarka:

1. **Kalbos dalių identifikavimas.** Panaudojant lietuvių kalbos morfologinį anotatorių, tekste atpažįstamos ir pažymimos kalbos dalys;
2. **Sintaksinis anotavimas.** Tekste atpažįstamos ir pažymimos sakinio dalys.
3. **Semantinis įvykių anotavimas.** Programa, panaudojant sporto įvykių atpažinimo taisykles, turi atpažinti sporto įvykius, jame dalyvaujančius dalyvius, vietą, laiką, rezultatą ir juos pažymėti atitinkamomis ontologijos klasėmis. Pirmiausia atpažįstamos daiktavardinės ir veiksmožodinės frazės, kurios sakinyje turi būti išsidėstę atitinkama tvarka.

Pasiekti rezultatai (taškai), sporto įvykius aprašančiuose tekstuose, paprastai rašomi skaičiais. Sakinyje, atpažinus įvykį, kuris reiškia laimėjimą arba pralaimėjimą, laimėjęs dalyvis susiejamas su didesne taškų suma, o pralaimėjęs – mažesne.

Automatinio sporto įvykių atpažinimo proceso diagrama pateikiama žemiau esančiame 2.1 paveiksle.



2.1 pav. Automatinis sporto įvykių atpažinimo procesas

Norit pritaikyti sporto įvykių atpažinimo taisyklės, tekstas pirmiausia turės būti išnagrinėtas gramatiškai. Tam galima panaudoti lietuvių kalbos morfologinį anotatorių, kuris automatiškai morfologiškai anotuoja tekstą, t.y. atpažįsta kalbos dalis ir jas pažymi. Kadangi sporto įvykių atpažinimo taisyklės bus paremtos tam tikrų kalbos dalių deriniais, automatiškai jų atpažinti nepavyktų, prieš tai neišsiaiškinus, kokios kalbos dalys yra naudojamos tekste.

Prieš anotuojant turės būti sudarytas kuo platesnis komandų pavadinimų, sportininkų, sporto organizacijų, trenerių, arenų pavadinimų rinkinys – žodynas. Tam, kad lyginant tekste naudojamus žodžius su įtrauktais į tam tikros srities žodyną, būtų galima iš karto atpažinti tam tikrus ontologijos konceptus ir taip tiksliau identifikuoti įvykusius sporto įvykius. Žodis, kuriame visos raidės yra didžiosios, gali būti laikomas organizacijos pavadinimu. Taip pat turi būti sudarytas žodžių/frazių, kurie dažniausiai naudojami, aprašant tam tikrus įvykius sporte, rinkinys.

3. SPORTO ĮVYKIŲ ONTOLOGIJOS REALIZACIJA

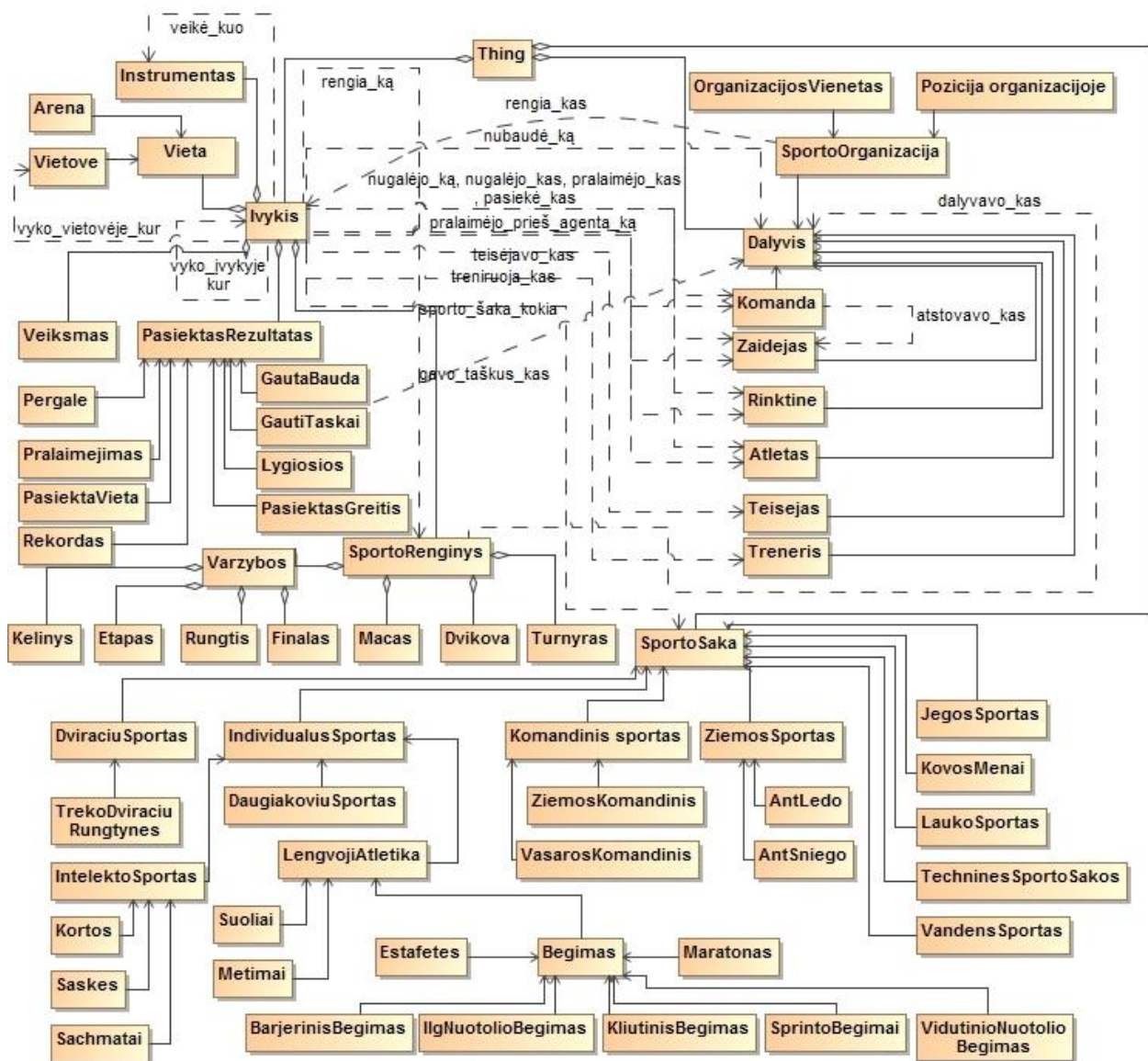
Sporto įvykių ontologija buvo sukurta Stanford University dėstytojų aprašyta metodika. Nustatant ontologijos apimtį ir sritį, buvo sudarytas klausimų sąrašas, į kuriuos turėjo atsakyti sporto įvykių ontologija. Sporto įvykių ontologijos pagrindas – įvykių ontologija (žr. *Bendroji įvykių ontologija*). Kiekvienas įvykis specifikuojamas kaip ontologijos klasė, turinti ryšius su susijusiais objektais, savybėmis, kitais įvykiais. Su sportu susiję konceptai ir jų hierarchija buvo sudaroma atsižvelgiant į BBC sporto ontologiją (žr. *BBC Sporto ontologija*). Sporto šakų hierarchija sudaryta atsižvelgiant į tai, kokias specifines savybes turi tam tikros klasės.

Klausimai į kuriuos turi atsakyti ontologija:

- Kas laimėjo rungtynes tarp komandos A ir komandos B?
- Kas laimėjo pasaulio futbolo čempionatą?
- Kas įmušė įvartį per varžybas X?
- Kiek žaidėjas Y įmetė taškų per rungtynes X?
- Kokį rekordą pasiekė sportininkė Y per čempionatą X?
- Kada įvyko įvykis?
- Kokios komandos kovojo varžybose X?
- Kokie atletai dalyvavo rungties X finale paskutinėse olimpinėse žaidynėse?
- Koks teisėjas teisėjavo rungtynėse X?
- Koku rezultatu laimėjo komanda A prieš komandą B?
- Kokį greitį pasiekė atletas rungtynėse X?

3.1. Pagrindinės ontologijos klasės

Žemiau esančiame 3.1 paveiksle pateikiama ontologijos klasių hierarchija. Klasės tarpusavyje susiję apibendrinimo (angl.: is-a) ir agregavimo (angl: kind-of) ryšiais. Agregavimo ryšys nurodo, iš kokių žemesnio lygio klasių susideda klasė. Pavyzdžiui klasė „Įvykis“ susideda iš klasių „Veiksmas“, „SportoRenginys“ ir „PasiektasRezultatas“. Apibendrinimo ryšys apibrėžia, kad žemesnio lygio klasės paveldi jų viršklasių savybes. Pavyzdžiui, klasės „ZiemosKomandinis“ ir „VasarosKomandinis“ yra klasės „KomandinisSportas“ poklasiai, todėl jie paveldi šios klasės savybes. Punktyrine linija pažymėtos klasių savybės.



3.1 pav. Sporto įvykių ontologija

Pagrindinės ontologijos klasės:

- *Thing* – OWL kalboje iš anksto apibrėžta klasė, kuri įtraukia viską. Visos kitos klasės yra šios klasės poklasiai.
- *Dalyvis* – Dalyvis, kuris dalyvauja įvykyje. Tai gali būti ir individualus sportininkas - atletas, komanda arba jos žaidėjas, treneris, teisėjas.
- *Ivykis* – Sporto įvykis, kuriame dalyvauja asmenys ir, kuriame pasiekiamas koks nors rezultatas.
- *PasiektasRezultatas* – įvykio sukurtas rezultatas. Tai gali būti: gauta bauda, gauti taškai, pasiektas greitis, pasiekta vieta, lygiosios, pergalė, pralaimėjimas, rekordas.
- *Vieta* – įvykio vyksmo vieta.
- *Instrumentas* – kuo pasinaudojant buvo įvykdytas įvykis.
- *Varzybos* – Sporto renginys, kuriame rungtiasi sportininkai, komandos. Jos gal būti suskirstytos į etapus arba sudarytos iš keleto kėlinių.

- *Etapas* – Sporto renginio dalis.
- *Kelinys* – Sporto rungtynių dalis.
- *Rungtynės* – Sportinės varžybos, kuriose kiekvienas dalyvis kovoja su kiekvienu.
- *Dvikova* – kai vienas sportininkas kovoja prieš kitą.
- *Macas* – kai viena komanda kovoja prieš kitą.
- *SportoOrganizacija* – Organizacija, organizuojanti sporto renginį.
- *SportoSaka* – Sporto šaka, susijusi su sporto įvykiu. Yra viršklasė klasėms: DviraciųSportas, IndividualusSportas, KomandinisSportas, ŽiemosSportas, JėgosSportas, KovosMenai, LaukoSportas, TechninesSportoSakos, VandensSportas.

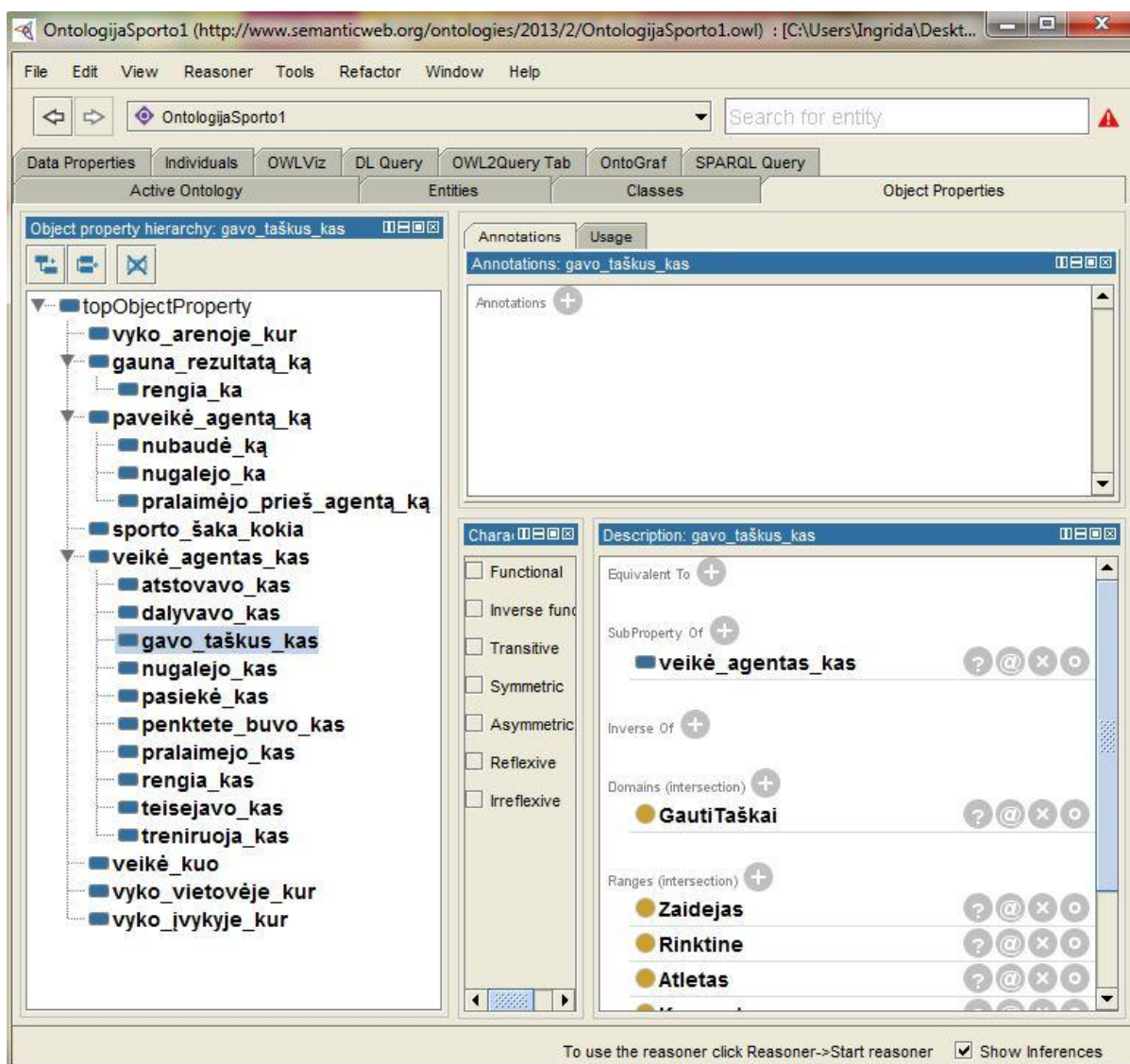
3.2.Ontologijos objektų savybės

Žemiau esančiame 3.2 paveiksle pateikiamas Protege langas, kuriame matomos ontologijos objektų savybės. Šios savybės parodo, kokius ryšius turi ontologijos klasės viena su kita.

Ontologijos objektų savybės:

- *vyko_arenoje_kur* – susieja įvykį su arena.
- *rengia_ką* – susieja įvykį su sporto renginiu.
- *nubaudė_ką* – susieja baudos ar baudos taškų gavimo įvykį su jį gavusiu dalyviu.
- *nugalėjo_ką* – nurodo, kuris įvykyje dalyvavęs dalyvis pralaimėjo.
- *pralaimėjo_prieš_agentą_ką* – nurodo, kuris įvykyje dalyvavęs dalyvis pralaimėjo prieš laimėjusį dalyvį.
- *sporto_šaka_kokia* – susieja įvykį su sporto šaka.
- *atstovavo_kas* – susieja komandą su ją atstovavusiu dalyviu.
- *dalyvavo_kas* – susieja sporto renginį su jame dalyvavusiu dalyviu.
- *gavo_taškus_kas* – susieja taškų gavimo įvykį su taškus gavusiu (surinkusiu) dalyviu.
- *nugalėjo_kas* – susieja sporto renginį su nugalėjusiu dalyviu.
- *pasiekė_kas* – susieja sporto renginį su jame ką nors pasiekusiu dalyviu.
- *penkete_dalyvavo_kas* – susieja įvykį su starto penkete dalyvavusiais komandos žaidėjais.
- *pralaimėjo_kas* – susieja įvykį su pralaimėjusiu dalyviu.
- *rengia_kas* – susieja įvykį su sporto organizacija.
- *teisėjavo_kas* – susieja sporto įvykį su jame dalyvavusiu teisėju.
- *treniruoja_kas* - susieja sporto įvykį su jame dalyvavusiu treneriu.
- *veikė_kuo* – susieja įvykį su instrumentu.
- *vyko_vietovėje_kur* – susieja įvykį ir vietovę.

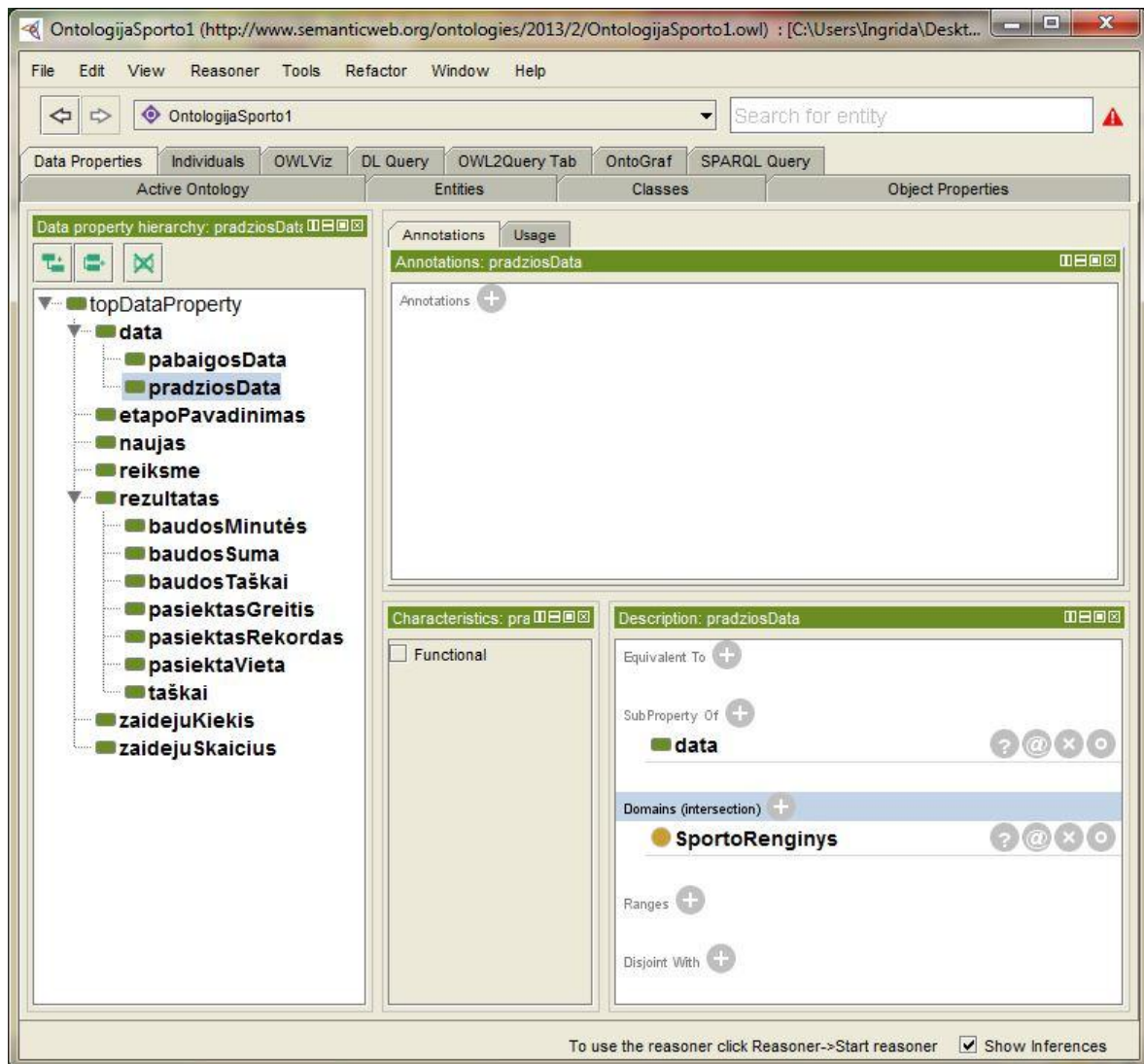
- *vyko_įvykyje_kur* – susieja įvykį su įvykiu, t.y. leidžia padalinti sudėtingą įvykį į keletą.



3.2 pav. Sporto įvykių ontologijos objektų savybės

3.3. Duomenų tipų savybės

Žemiau esančiame 3.3 paveiksle pateikiamas Protege langas, kuriame matomos ontologijos duomenų tipų savybės. Šios savybės skirtos susieti ontologijos elementus su duomenų tipu.



3.3 pav. Sporto įvykių ontologijos duomenų tipų savybės

Pagrindinės ontologijos duomenų tipų savybės:

- *data* – susieja sporto renginį su datos tipu, t.y. nurodo, kad renginiui galima įvesti datą.
- *rezultatas* – skirtas susieti rezultatų gavimo įvykius, su kiekvienam įvykiui tinkamu duomenų tipu.

4. ĮVYKIO ATPAŽINIMO TEKSTE METODIKA

4.1. Įvykio atpažinimo taisyklės

Žemiau pateikiamos taisyklės, pagal kurias galima atpažinti įvykius, juose dalyvaujančius dalyvius, jų surinktus rezultatus. Įvykio esybės atpažįstamos pagal taisykles, šablonus, įvertinus kolokacijas ir atlikus paiešką žodyne, jeigu prieš tai tokia esybė jau buvo užfiksuota.

Taisyklės pateikiamos tokiu principu (ne visi punktai privalomi):

- Esybės pavadinimas;
- Atpažinimo sąlygos;
- Atpažinimo būdai;
- Taisyklės aprašas natūralia kalba;
- Formalus taisyklės aprašas, panaudojant aibių teoriją ir pirmos eilės predikatų logiką;
- Esybės atpažinimo proceso veiklos diagrama.

Jeigu atpažįstama pagal šablonus, pateikiama lentelė, kurioje nurodoma:

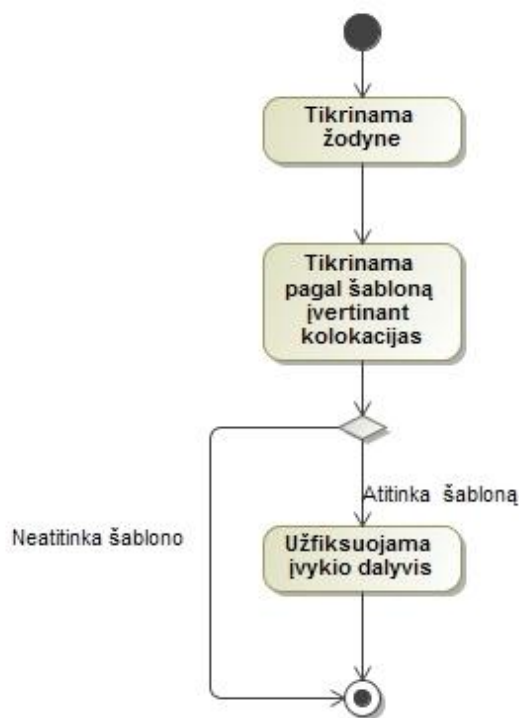
- Šablonas;
- Papildomos sąlygos su morfologinėmis savybėmis, kurias jis turi atitikti;
- Pavyzdinis tekstas su pažymėta pagal šabloną atpažinta esybe;
- Po lentele pateikiamas kolokacinių žodžių sąrašas. Kolokacija yra dažnai vartojamų žodžių junginys, seka.

Esybės atpažinimo būdai gali būti tokie:

- Tikrinimas žodyne. Tikrinama, ar žodis, ar žodžių rinkinys nėra įtrauktas į, iš anksto tai esybei sudarytą, žodyną.
- Pagal šabloną. Ieškoma ar sakinyje nėra frazių, atitinkančių nurodytą šabloną. Šablonas – reguliari išraiška, kuri dažnai naudojama įvardijant tam tikros srities esybes. Įvertinamos kolokacijos su kolokaciniais žodžiais, iš prieš tai taisyklei sudaryto, kolokacinių žodžių sąrašo.

1. Komandos atpažinimas

Komandos atpažinimo procesas patiekiamas 4.2 paveiksle. Pagal jį aprašytos sporto komandos atpažinimo taisyklės, pateikiamos po proceso diagrama.



4.1 pav. Komandos atpažinimo procesas

Komandos, dalyvio, sporto renginio atpažinimui nuspręsta panaudoti tuos pačius principus, kaip ir įvardintų esybių atpažinimui (NER), tai: atpažinimas pagal žodyną, reguliarios išraiškos atitikimas [25].

Komandos atpažinimo žingsniai:

- Patikrinama žodyne t.y. tekste esantys žodžiai palyginami su žodžiais įtrauktais į komandų pavadinimų žodyną.
- Patikrinama pagal šablonus t.y. patikrinant ar tekste esantys žodžių rinkiniai atitinka kurį nors iš žemiau esančioje 4.1 lentelėje patiktų šablonų. Šablonas „[a-zA-Z]“ reiškia, kad tarp kabučių esantį komandos pavadinimą gali sudaryti visos abėcėlės didžiosios ir mažosios raidės. Šablonas [A-Z] nurodo, kad komandos pavadinimas gali būti tik iš visų abėcėlės raidžių, kurios turi būti didžiosios. [Miestas] – nurodo, kad prieš komandos pavadinimą gali būti įvardintas bet koks miestas. Kartais sporto komandos pavadinimo šabloną gali teksto fragmentas, neturintis nieko bendro su sportu, todėl reikia įvertinti aplink jį esančių žodžių išsidėstymą t.y. kolokacijas.

4.1 lentelė. Šablonai skirti komandos atpažinimui

Šablonas su kolokaciniais žodžiais	Papildomos sąlygos	Pavyzdys
[Miestas] „[a-zA-Z]“ klubas	Kolokacinis žodis turi būti daugiskaitos galininko, kilmininko arba vardininko	„Keturioliktą vietą Rusijos pirmo futbolo diviziono čempionate užimantis dviejų lietuvių atstovaujamas Maskvos „Torpedo“ klubas kontroliniame

	linksnyje ([dkt., dgs., V.]; [dkt., dgs., G.]; [dkt., dgs., K.]) Miesto pavadinimas turi būti	mače 2:1 (1:0) nugalėjo šalies aukščiausiosios lygos lyderius Maskvos CSKA futbolininkus.“ [30].
[Miestas] [A-Z] [sporto šakos atstovo pavadinimas]	kilmininko linksnyje.	„Keturioliktą vietą Rusijos pirmo futbolo diviziono čempionate užimantis dviejų lietuvių atstovaujamas Maskvos „Torpedo“ klubas kontroliniame mače 2:1 (1:0) nugalėjo šalies aukščiausiosios lygos lyderius Maskvos CSKA futbolininkus.“ [30].
[A-Z] komanda	Kolokacinis žodis turi būti galininko, kilmininko arba vardininko linksnyje.	„Nors viešoji įstaiga „LKKA futbolo ir rankinio sporto centras“ įkurta 2005 m. rugsėjo 2 d., tačiau LKKA komanda žaidė jau 2004 m. sezone <...>.“ [31].
„[a-zA-Z]“ komanda		„Po trijų kėlinių rezultatu 57:55 pirmavo „Kumščio“ komandos krepšininkai.“ [32].
Komanda „[a-zA-Z]“		Komanda “AB Utenos alus”.

- Įvertinamos, pagal šabloną atpažintų, teksto dalių kolokacijos su žodžiais: komanda, klubas, ekipa, krepšininkė, futbolininkas, vienuolikė, futbolininkė, tinklininkas, tinklininkė, rankininkas, rankininkė, beisbolininkas, beisbolininkė, dažasvydininkas, dažasvydininkė, vandensvydininkas, vandensvydininkė, irklutojas, irklutoja, buriuotojas, buriuotoja, lenktynininkas, lenktynininkė.
- Patikrinant pagal žemiau esančią taisyklę, užfiksuojamas įvykio dalyvis, Apibrėžiama, kad s yra sakinių aibės elementas, o d yra dalyvių aibės elementas. Apibrėžiama, kad $s \in S$, čia S – sakinių aibė.

Atpažinimo rezultatas – $d \in D$, čia D – dalyvių aibė

Jeigu kiekvienas s priklauso sakinių aibei, kiekvienas d priklauso dalyvių aibei, tai sakinyje fiksuojamas dalyvis (1).

$$\forall s \in S \quad \forall d \in D \Rightarrow [\text{priklauso}(d, s)] \quad (1)$$

2. Dalyvio atpažinimas

Dalyvio atpažinimo procesas yra toks pats kaip ir komandos (3.2 paveikslas). Pagal jį aprašytos dalyvio atpažinimo taisyklės, pateikiamos žemiau.

Dalyvio atpažinimo žingsniai:

- Patikrinama žodyne t.y. tekste esantys žodžiai palyginami su žodžiais įtrauktais į dalyvių pavadinimų žodyną.
- Patikrinama pagal šablonus t.y. patikrinant ar tekste esantys žodžių rinkiniai atitinka kurį nors iš žemiau esančioje 3.2 lentelėje patiktų šablonų. Šablonas „[a-zA-Z]“ reiškia, kad tarp kabučių esantį dalyvio pavadinimą gali sudaryti visos abėcėlės didžiosios ir mažosios raidės.

4.2 lentelė. Šablonai skirti dalyvio atpažinimui

Šablonas su kolokaciniais žodžiais	Papildomos sąlygos	Pavyzdys
[sporto šakos atstovo pavadinimas] „[a-zA-Z]“	Kolokacinis žodis turi vardininko arba galininko linksnyje.	„Tenisininkas Lukas Mugevičius baigė pasirodymą prestižinio „Orange Bowl“ turnyro vienetų varžybose.“[33].
[Pilietybė] [a-zA-Z]	Pilietybės pavadinimas turi vardininko, galininko arba kilmininko linksnyje.	„Vienetų varžybų nugalėtoju pirmą sykį tapo favoritu laikytas pajėgiausias planetos tenisininkas serbas Novakas Džokovičius.“[34].
raketė [a-zA-Z]		„Pirma pasaulio raketė Novakas Džokovičius 6:3, 7:6(4) nugalėjo septintą pasaulio raketę Juaną Martiną del Potro ir pateko į ATP 500 serijos „Dubai Duty Free Tennis Championships“ finalą. Mačas truko valandą ir 45 minutes.“.[35]

- Įvertinamos, pagal šabloną atpažintų, teksto dalių kolokacijos su žodžiais: žaidėjas, žaidėja, atletas, atletė, raketė, krepšininkas, krepšininkė, futbolininkas, futbolininkė, tinklininkas, tinklininkė, rankininkas, rankininkė, dviratininkas, dviratininkė, tenisininkas, tenisininkė, daugiakovininkas, daugiakovininkė, šachmatininkas, šachmatininkė, bėgikas, bėgikė, maratonininkas, maratonininkė, sprinteris, sprinterė, metikas, metikė, penkiakovininkas, penkiakovininkė, septynkovininkas, septynkovininkė, kultūristas, kultūristė, beisbolininkas, beisbolininkė, dažasvydininkas, dažasvydininkė, vandensvydininkas, vandensvydininkė, riedutininkas, riedutininkė, boksiniškas, boksiniškė, kovotojas, kovotoja, imtynininkas, imtynininkė, karatistas, karatistė, žvejys, žvejė, parašiutininkas, parašiuntininkė, lenktynininkas, lenktynininkė, motociklininkas, motociklininkė, irkluojuojas, irkluojuoja, buriuojuojas, buriuojuoja, plaukikas, plaukikė.
- Patikrinant pagal žemiau esančią taisyklę, užfiksuojamas įvykio dalyvis.
Apibrėžiama, kad *s* yra sakinių aibės elementas, o *d* yra dalyvių aibės elementas.

Apibrėžiama, kad $s \in S$, čia S – sakinių aibė.

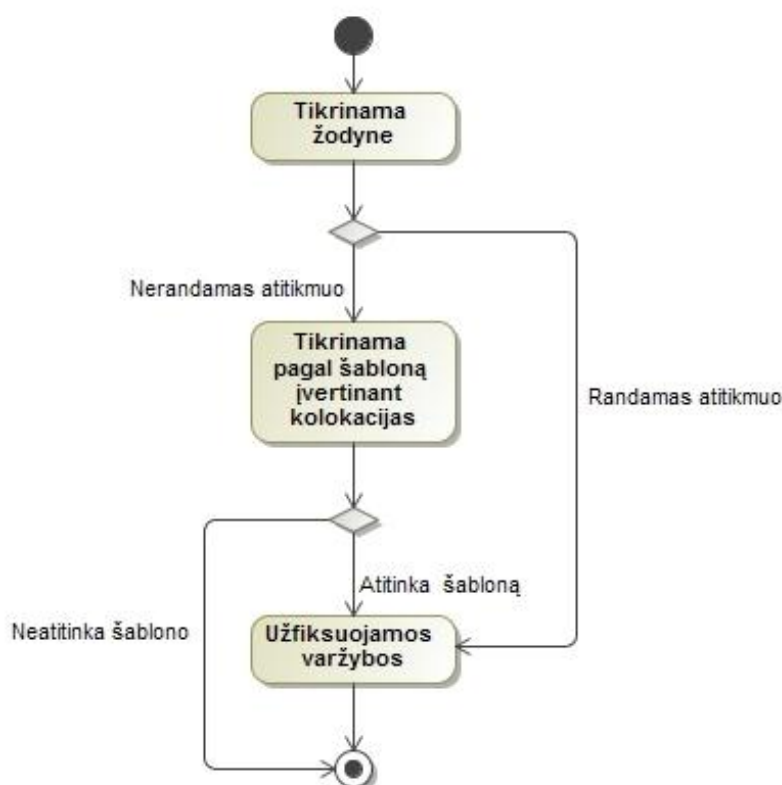
Atpažinimo rezultatas – $d \in D$, čia D – dalyvių aibė

Jeigu kiekvienas s priklauso sakinių aibei, kiekvienas d priklauso dalyvių aibei, tai sakinyje fiksuojamas dalyvis (2).

$$\forall s \in S \forall d \in D \Rightarrow [\text{priklauso}(d, s)] \quad (2)$$

3. Sporto renginio atpažinimas

Sporto renginio atpažinimo procesas pateikiamas 4.3 paveiksle. Pagal jį aprašytos sporto renginio atpažinimo taisyklės, pateikiamos po proceso diagrama.



4.2 pav. Sporto renginio atpažinimo procesas

Sporto renginio atpažinimo žingsniai:

- Patikrinama žodyne t.y. tekste esantys žodžiai palyginami su žodžiais įtrauktais į sporto renginių pavadinimų žodyną.
- Patikrinama pagal šablonus t.y. patikrinant ar tekste esantys žodžių rinkiniai atitinka kurį nors iš žemiau esančioje 4.3 lentelėje patiktų šablonų. Šablonas [A-Z-0-9] reiškia, kad sporto renginio etapo pavadinimą pavadinimą gali sudaryti visos abėcėlės didžiosios raidės ir po jų sekantys visi sveikieji skaičiai.

4.3 lentelė. Šablonai skirti sporto renginio atpažinimui

Šablonas su kolokaciniais žodžiais	Papildomos sąlygos	Pavyzdys
„[a-zA-Z]“ taurės varžybos	Kolokacinis žodis turi galininko linksnyje.	„Šiomet atnaujintas Lietuvos šalės futbolo asociacijos (LSFA) „Select“ taurės varžybas laimėjo Kazlų Rūdos „STIHL SC“ komanda.“ [36].
[A-Z-0-9] etapas [A-Z] grupės rungtynės	Kolokacinis žodis turi vietininko linksnyje.	„Ketvirtadienį Kauno „Žalgirio“ komanda užbaigė savo pasirodymą šių metų Eurolygos sezono - paskutinėse TOP-16 etapo E grupės rungtynėse žalgiriečiai namuose 90:81 (26:25, 24:20, 16:18, 24:18) nugalėjo Vokietijos čempionus Bambergo „Brose Baskets“ krepšininkus.“[37]
[0-9] [Šalis] rungtynės		„2012-2013 metų sezono Prancūzijos moterų krepšinio taurės turnyro pusfinalio rungtynėse Giedrės Paugaitės atstovaujamas Mondvilio USO klubas namuose 64:79 (18:20, 15:22, 22:26, 11:11) nusileido Sent Herbleno „Nantes Reze“ krepšininkams ir nepateko į finalą.“[38].
[Šalis] rungtynės		„Devynioliktą pralaimėjimą Ispanijos vyrų krepšinio lygos XXX turo rungtynėse patyrė bei į šešioliktą vietą smuktelėjo Vytenio Čižausko Valjadolido "Blancos de Rueda" klubas.“[39]
[A-Z] „[a-zA-Z]“ turnyras		„JAV, Kalifornijos valstijoje, vykstančio Teniso profesionalų asociacijos (ATP) "World Tour Masters 1000" serijos turnyro "BNP Paribas Open", kurio prizų fondą sudaro 4,33 mln. JAV dolerių, vienetų atrankos varžybų pirmajame rate antradienį lietuvis Ričardas Berankis 7:6 (7-2), 6:1 įveikė 133-iają pasaulio raketę 21-erių metų amerikietį Rhyne'ą Williamsą.“[40]
[0-9] [Miestas] maratonas	Kolokacinis žodis turi galininko linksnyje.	„28-ąjį Hamburgo maratoną šiandien laimėjo Diana Lobačevskė - 2 val. 29 min.17 sek.“ [41]

- Įvertinamos, pagal šabloną atpažintų, teksto dalių kolokacijos su žodžiais: turnyras, rungtynės, grupės rungtynės etapas, varžybos, mačas, kėlinys, sezonas, finalas, pusfinalis, ketvirtfinalis, lyga, turas, olimpiada sporto šakų pavadinimai (sporto šakų pavadinimai yra įtraukti į ontologiją).
- Patikrinus pagal žemiau esančią taisyklę, užfiksuojamas sporto renginys.

Atpažinimo rezultatas – $v \in V$, čia V – varžybų aibė

Jeigu kiekvienas s priklauso sakinių aibei, kiekvienas v priklauso varžybų aibei, tai sakinyje fiksuojamos varžybos (3).

$$\forall s \in S \quad \forall v \in V \Rightarrow [\text{priklauso}(v, s)] \quad (3)$$

4. Varžybų dalyvio atpažinimas

Varžybų dalyvio atpažinimo procesas pateikiamas 4.4 paveiksle. Pagal jį aprašytos sporto varžybų dalyvio atpažinimo taisyklės, pateikiamos po proceso diagrama.



4.3 pav. Varžybų dalyvio atpažinimo procesas

Norint užfiksuoti, kad dalyvis yra varžybų dalyvis, pirmiausia, sakinyje turi būti atpažintas dalyvis ir varžybos.

Jei kiekvienas s priklauso sakinių aibei, kiekvienas d priklauso dalyvių aibei, kiekvienas v priklauso varžybų aibei, yra užfiksuota, kad dalyvis d priklauso sakiniui s ir varžybos v priklauso sakiniui s , tai varžybose v dalyvavo dalyvis d (4).

$$\forall s \in S \quad \forall d \in D \quad \forall v \in V [\text{priklauso}(d, s) \wedge \text{priklauso}(v, s)] \Rightarrow [\text{varžybų_dalyvis}(v, d)] \quad (4)$$

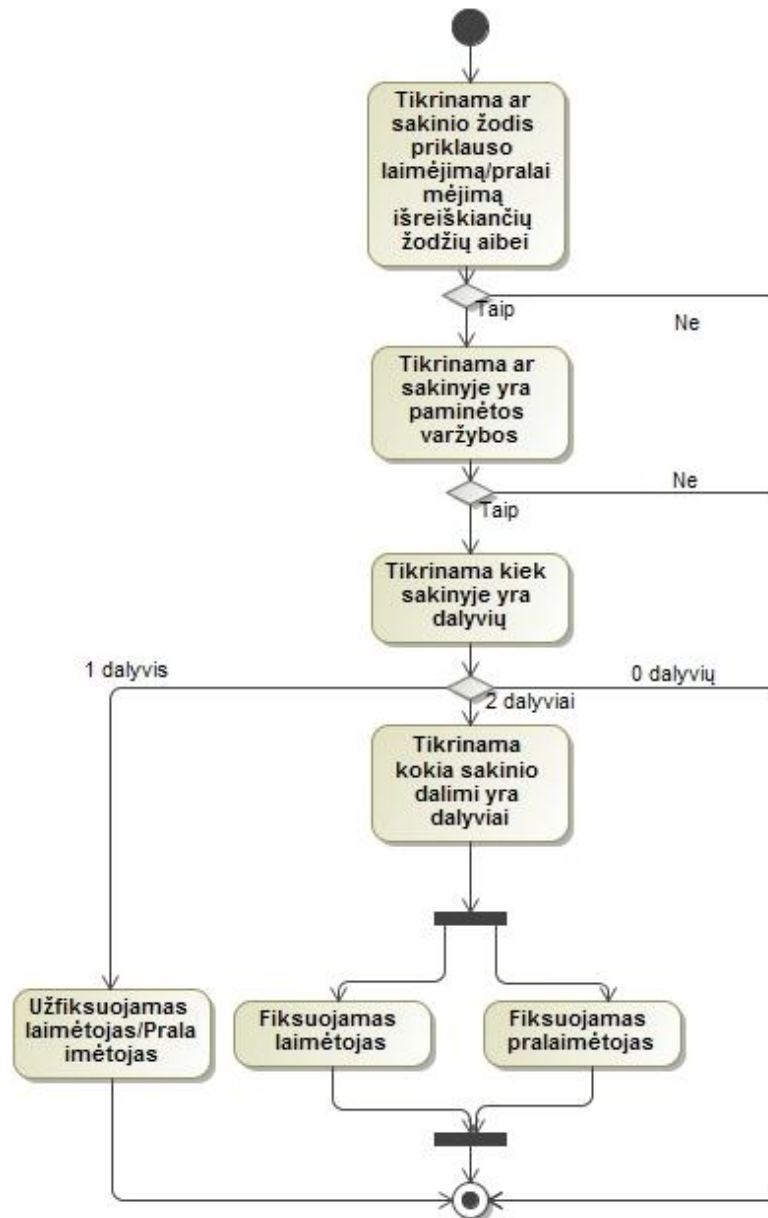
Jei kiekvienas s priklauso sakinių aibei, kiekvienas d priklauso dalyvių aibei, kiekvienas v priklauso varžybų aibei, yra užfiksuota, kad dalyvis d priklauso sakiniui s ir varžybos v priklauso sakiniui s , tai dalyvis d yra varžybų v dalyvis (5).

$$\forall s \in S \quad \forall d \in D \quad \forall v \in V [\text{priklauso}(d, s) \wedge \text{priklauso}(v, s)] \Rightarrow [\text{dalyvavo_kur}(d, v)] \quad (5)$$

Varžybų dalyvio atpažinimo procesas pateikiamas 3.4 paveiksle.

5. Laimėjimo (pralaimėjimo) įvykio atpažinimas

Laimėjimo/pralaimėjimo atpažinimo procesas pateikiamas 4.5 paveiksle. Pagal jį aprašytos laimėjimo/ pralaimėjimo atpažinimo taisyklės, pateikiamos po proceso diagrama.



4.4 pav. Laimėjimo/ pralaimėjimo atpažinimas

Laimėjimo įvykio atpažinimas vyksta, pirmiausia, patikrinant laimėjimą, apibrėžiančių raktinių žodžių žodyne (laimėjo, nugalėjo, įveikė, nugalėtojo titulą apgynė, palaužė, pergale iškovojo, pergale išplėšė, pranoko, priverstė pasiduoti, triumfavo).

Tada tikrinama, ar, laimėjimą apibrėžiantys žodžiai ir dalyviai, atitinka žemiau nurodytą daiktavardinių ir veiksmažodinių frazių seką ir, ar turi nurodytus morfologinius požymius.

Jeigu yra du dalyviai, turi būti tokia frazių seka:

- [Daikt f.] + [Veiks.f] + [Daikt f.]

Frazių morfologiniai požymiai:

- Daikt f.– dalyvis arba šalia dalyvio einantis kolokacinis žodis vardininko linksnyje;

- Veiks.f – laimėjimą/pralaimėjimą išreiškiantis žodis turi būti veiksmažodis, nesangražinis, tiesioginė nuosaka, būtojo k. laiko, 3 amenuotės.
- Daikt f.– antras dalyvis arba šalia dalyvio einantis kolokacinis žodis galininko linksnyje;

Jeigu yra vienas dalyvis, turi būti tokia frazių seka:

- [Daikt f.] + [Veiks.f] + [Daikt f.]

Frazių morfologiniai požymiai:

- Daikt f.– varžybų pavadinimas galininko linksnyje;
- Veiks.f – laimėjimą/pralaimėjimą išreiškiantis žodis turi būti veiksmažodis, nesangražinis, tiesioginė nuosaka, būtojo k. laiko, 3 asmenuotės.
- Daikt f.– dalyvis arba šalia dalyvio einantis kolokacinis žodis vardininko linksnyje;

Atpažinimo rezultatas – $l \in L$, čia L – laimėjimą išreiškiančių žodžių aibė

Jeigu kiekvienas s priklauso sakinių aibei, kiekvienas l priklauso laimėjimų aibei, tai sakinyje fiksuojamas laimėjimas (6).

$$\forall s \in S \forall l \in L \Rightarrow [\text{priklauso}(l, s)] \quad (6)$$

Jei kiekvienas s priklauso sakinių aibei, kiekvienas d priklauso dalyvių aibei, kiekvienas v priklauso varžybų aibei, kiekvienas, l priklauso laimėjimų aibei, yra užfiksuota, kad dalyvis d priklauso sakiniui s , varžybos v priklauso sakiniui s , sakinyje dalyvis yra veiksnys ir sakinyje laimėjimas yra tarinys, tai varžybas v laimėjo dalyvis d (7).

$$\forall s \in S \forall d \in D \forall l \in L \forall v \in V [\text{priklauso}(d, s) \wedge \text{priklauso}(l, s) \wedge \text{priklauso}(v, s) \wedge \text{veiksnys}(s, d) \wedge \text{tarinys}(s, l)] \Rightarrow [\text{laimėjo}(v, d)] \quad (7)$$

Jei kiekvienas s priklauso sakinių aibei, kiekvienas d priklauso dalyvių aibei, kiekvienas v priklauso varžybų aibei, kiekvienas, l priklauso laimėjimų aibei, yra užfiksuota, kad dalyvis d priklauso sakiniui s , varžybos v priklauso sakiniui s , sakinyje dalyvis nėra veiksnys ir sakinyje laimėjimas yra tarinys, tai varžybas v pralaimėjo dalyvis d (8).

$$\forall s \in S \forall d \in D \forall l \in L \forall v \in V [\text{priklauso}(d, s) \wedge \text{priklauso}(l, s) \wedge \text{priklauso}(v, s) \wedge \neg \text{veiksnys}(s, d) \wedge \text{tarinys}(s, l)] \Rightarrow [\text{pralaimėjo}(v, d)] \quad (8)$$

Pralaimėjimo įvykio atpažinimas vyksta patikrinant laimėjimą apibrėžiančių raktinių žodžių žodyne (pralaimėjo, pasidavė, nusileido).

Tada tikrinama, ar, pralaimėjimą apibrėžiantys žodžiai ir dalyviai, atitinka žemiau nurodytą daiktavardinių ir veiksmažodinių frazių seką ir ar turi nurodytus morfologinius požymius.

Jeigu yra du dalyviai, turi būti tokia frazių seka:

- [Daikt f.] + [Veiks.f] + [Daikt f.]

Frazių morfologiniai požymiai:

- Daikt f.– dalyvis arba šalia dalyvio einantis kolokacinis žodis vardininko linksnyje;
- Veiks.f – laimėjimą/pralaimėjimą išreiškiantis žodis turi būti veiksmažodis, nesangražinis, tiesioginė nuosaka, būtojo k. laiko, 3 asmenuotės.
- Daikt f.– antras dalyvis arba šalia dalyvio einantis kolokacinis žodis naudininko linksnyje;

Atpažinimo rezultatas – $p \in P$, čia P – pralaimėjimą išreiškiančių žodžių aibė

Jeigu kiekvienas s priklauso sakinių aibei, kiekvienas p priklauso pralaimėjimų aibei, tai sakinyje fiksuojamas pralaimėjimas (9).

$$\forall s \in S \forall p \in P \Rightarrow [\text{priklauso}(p, s)] \quad (9)$$

Jei kiekvienas s priklauso sakinių aibei, kiekvienas d priklauso dalyvių aibei, kiekvienas v priklauso varžybų aibei, kiekvienas p priklauso pralaimėjimų aibei, yra užfiksuota, kad dalyvis d priklauso sakiniui s , varžybos v priklauso sakiniui s , sakinyje dalyvis yra veiksnys ir sakinyje pralaimėjimas yra tarinys, tai varžybas v pralaimėjo dalyvis d (10).

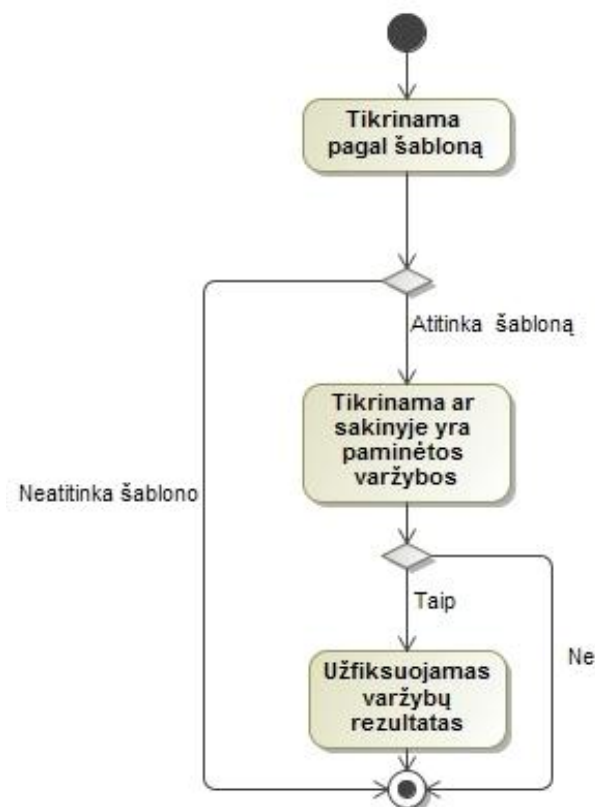
$$\forall s \in S \forall d \in D \forall p \in P \forall v \in V [\text{priklauso}(d, s) \wedge \text{priklauso}(p, s) \wedge \text{priklauso}(v, s) \wedge \text{veiksnys}(s, d) \wedge \text{tarinys}(s, p)] \Rightarrow [\text{pralaimėjo}(v, d)] \quad (10)$$

Jei kiekvienas s priklauso sakinių aibei, kiekvienas d priklauso dalyvių aibei, kiekvienas v priklauso varžybų aibei, kiekvienas p priklauso pralaimėjimų aibei, yra užfiksuota, kad dalyvis d priklauso sakiniui s , varžybos v priklauso sakiniui s , sakinyje dalyvis nėra veiksnys ir sakinyje laimėjimas yra tarinys, tai varžybas v laimėjo dalyvis d (11).

$$\forall s \in S \forall d \in D \forall p \in P \forall v \in V [\text{priklauso}(d, s) \wedge \text{priklauso}(p, s) \wedge \text{priklauso}(v, s) \wedge \neg \text{veiksnys}(s, d) \wedge \text{tarinys}(s, p)] \Rightarrow [\text{laimėjo}(v, d)] \quad (11)$$

6. Rezultato atpažinimas

Varžybų rezultato atpažinimo procesas patiekiamas 4.6 paveiksle. Pagal jį aprašytos varžybų rezultato atpažinimo taisyklės, pateikiamos po proceso diagrama.



4.5 pav. Varžybų rezultato atpažinimo procesas

Varžybų rezultato atpažinimo žingsniai:

- Patikrinama pagal šablonus t.y. patikrinant ar tekste esantys simbolių rinkiniai atitinka kurį nors iš žemiau esančioje 4.4 lentelėje patiktų šablonų. Pagal šabloną atpažįstamas galutinis varžybų rezultatas yra pažymėtas raudonai.

4.4 lentelė. Šablonai skirti rezultato atpažinimui

Šablonas	Pavyzdys
[0-9]:[0-9]	2:1
[0-9]:[0-9] ([0-9]:[0-9])	2:1 (0:1)
[0-9]:[0-9] ([0-9]:[0-9], [0-9]:[0-9], [0-9]:[0-9], [0-9]:[0-9]) [Įvykis]	Mondvilio USO klubas namuose 64:79 (18:20, 15:22, 22:26, 11:11) nusileido Sent Herbleno „Nantes Reze“.
Rezultatas [0-9]:[0-9] ([0-9]:[0-9], [0-9]:[0-9], [0-9]:[0-9], [0-9]:[0-9])	įtikinamai nugalėję varžovus rezultatu 81:60 (18:18, 15:18, 26:10, 22:14).
[0-9]:[0-9] ([0-9]:[0-9], [0-9]:[0-9], [0-9]:[0-9], [0-9]:[0-9])	Maskvos „Torpedo“ klubas kontroliniame mače 2:1 (1:0) nugalėjo.

- Įvertinamos, pagal šabloną atpažintų, teksto dalių kolokacijos su žodžiais: taškai, rezultatas.
- Patikrinus pagal žemiau esančią taisyklę, užfiksuojamas varžybų rezultatas.

Atpažinimo rezultatas: $x1, x2 \in X$, čia X – galutinius rezultatus išreiškiantys skaičiai

Jei kiekvienas s priklauso sakinių aibei, kiekvienas x priklauso rezultatų išreiškiančių skaičių aibei, yra užfiksuota, kad rezultatą išreiškiantis skaičius x priklauso sakiniui s , tai egzistuoja bent vienas rezultatas r , kuris turi rezultato reikšmę x ir priklauso sakiniui (12).

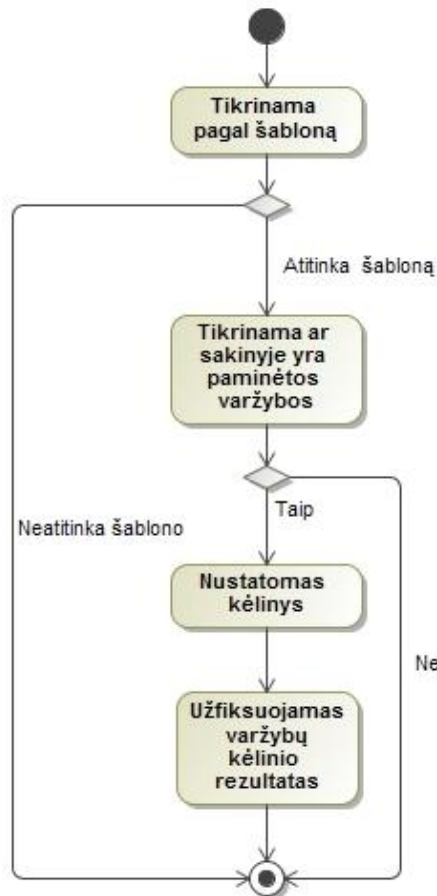
$$\forall s \in S \forall x \in X [\text{priklauso}(x, s)] \Rightarrow \exists r \in R [\text{rezultato_reikšmė}(r, x) \wedge \text{priklauso}(r, s)]; \quad (12)$$

čia R – rezultatų aibė.

Jei kiekvienas s priklauso sakinių aibei, kiekvienas r priklauso rezultatų aibei, kiekvienas v priklauso varžybų aibei, yra užfiksuota, kad varžybos v priklauso sakiniui s ir rezultatas r priklauso sakiniui s , tai vai varžybų v rezultatas yra r (13).

$$\forall s \in S \forall r \in R \forall v \in V [\text{priklauso}(v, s) \wedge \text{priklauso}(r, s)] \Rightarrow [\text{varžybų_rezultatas}(v, r)] \quad (13)$$

Varžybų kėlinio rezultato atpažinimo procesas pateikiamas 4.7 paveiksle. Pagal jį aprašytos varžybų kėlinio rezultato atpažinimo taisyklės, pateikiamos po proceso diagrama.



4.6 pav. Varžybų kėlinio rezultato atpažinimo procesas

Varžybų kėlinio rezultato atpažinimo žingsniai:

- Patikrinama pagal šablonus t.y. patikrinant ar tekste esantys simbolių rinkiniai atitinka kuri nors iš žemiau esančioje 4.5 lentelėje patiktų šablonų. Pagal šabloną atpažįstamas varžybų kėlinio, raundo yra pažymėtas raudonai.

4.5 lentelė. Šablonai skirti kėlinio rezultato atpažinimui

Šablonas	Pavyzdys
[0-9]:[0-9] ([0-9]:[0-9])	2:1 (0:1)
[0-9]:[0-9] ([0-9]:[0-9], [0-9]:[0-9], [0-9]:[0-9], [0-9]:[0-9])	Maskvos „Torpedo“ klubas kontroliniame mače 2:1 (1:0) nugalėjo
[0-9]:[0-9] ([0-9]:[0-9], [0-9]:[0-9], [0-9]:[0-9], [0-9]:[0-9]) [Įvykis]	Mondvilio USO klubas namuose 64:79 (18:20, 15:22, 22:26, 11:11) nusileido Sent Herbleno „Nantes Reze“
Rezultatas [0-9]:[0-9] ([0-9]:[0-9], [0-9]:[0-9], [0-9]:[0-9], [0-9]:[0-9])	įtikinamai nugalėję varžovus rezultatu 81:60 (18:18, 15:18, 26:10, 22:14).

- Įvertinamos, pagal šabloną atpažintų, teksto dalių kolokacijos su žodžiais: taškai, rezultatas.
- Patikrinus pagal žemiau esančią taisyklę, užfiksuojamas varžybų kėlinio rezultatas.

Atpažinimo rezultatas: $y1, y2 \in Y$, čia Y – kėlinio/raundo rezultatus išreiškiantys skaičiai, $k \in K$, čia K – kėlinių/raundų aibė.

Jei kiekvienas s priklauso sakinių aibei, kiekvienas y priklauso kėlinio/raundo rezultatus išreiškiančių skaičių aibei, kiekvienas k priklauso kėlinių aibei, yra užfiksuota, kad kėlinio/raundo rezultatą išreiškiantis skaičius y priklauso sakiniui s ir kėlinys k priklauso sakiniui s , tai egzistuoja bent vienas kėlinio rezultatas kr , kuris turi kėlinio/raundo rezultato reikšmę y ir priklauso sakiniui (14).

$$\forall s \in S \forall y \in Y \forall k \in K [\text{priklauso}(y, s) \wedge \text{priklauso}(k, s)] \Rightarrow \exists kr \in KR$$

$$[\text{kėlinio_rezultato_reikšmė}(kr, y) \wedge \text{priklauso}(kr, s)]; \quad (14)$$

čia KR – kėlinio rezultatų aibė.

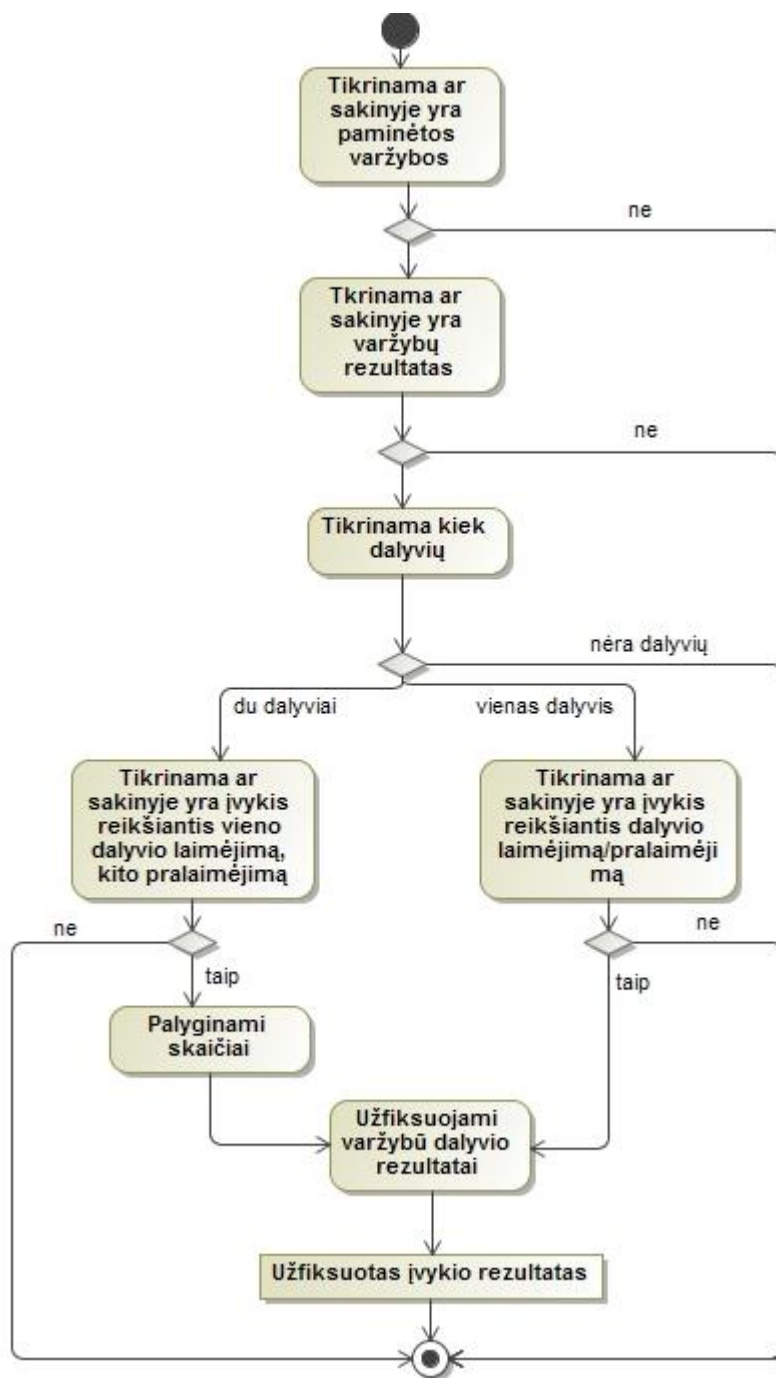
Jei kiekvienas s priklauso sakinių aibei, kiekvienas kr priklauso kėlinio rezultatų aibei, kiekvienas v priklauso varžybų aibei, kiekvienas k priklauso kėlinių aibei, yra užfiksuota, kad varžybos v priklauso sakiniui s , kėlinio rezultatas kr priklauso sakiniui s ir kėlinys priklauso sakiniui s , tai vai varžybų v kėlinio k rezultatas yra r (15).

$$\forall s \in S \forall kr \in KR \forall v \in V \forall k \in K [\text{priklauso}(v, s) \wedge \text{priklauso}(kr, s) \wedge \text{priklauso}(k, s) \wedge]$$

$$\Rightarrow [\text{varžybų_kėlinio_rezultatas}(v, k, kr)] \quad (15)$$

7. Varžybų dalyvio rezultato atpažinimas

Varžybų dalyvio rezultato atpažinimo procesas patiekiamas 4.8 paveiksle. Pagal jį aprašytos varžybų dalyvio rezultato atpažinimo taisyklės, pateikiamos po proceso diagrama.



4.7 pav. Dalyvio rezultato atpažinimo procesas

Bendro rezultato atpažinimas, kai du dalyviai:

Jei kiekvienas s priklauso sakinių aibei, dalyviai $d1$ ir $d2$ priklauso dalyvių aibei, kiekvienas v priklauso varžybų aibei, rezultatai $r1, r2$ priklauso rezultatų aibei, yra užfiksuota, kad sakiniui priklauso dalyviai $d1$ ir $d2$, varžybos priklauso sakiniui, rezultatai $r1, r2$ priklauso sakiniui, varžybas v laimėjo dalyvis $d1$ arba pralaimėjo dalyvis $d2$, rezultatas $r1$ didesnis už rezultatą $r2$, tai dalyvio $d1$ rezultatas yra $r1$ ir dalyvio $d2$ rezultatas yra $r2$ (16).

$$\forall s \in S \forall d1, d2 \in D \forall v \in V \forall r1, r2 \in R [\text{priklauso}(s, d1) \wedge \text{priklauso}(s, d2) \wedge \text{priklauso}(v, s) \wedge \text{priklauso}(r1, s) \wedge \text{priklauso}(r2, s) \wedge [\text{laimėjo}(v, d1) \vee \text{pralaimėjo}(v, d2)]] \wedge \quad (16)$$

$$\text{didesnis_už}(r1, r2) \Rightarrow [\text{dalyvio_rezultatas}(d1, r1) \wedge \text{dalyvio_rezultatas}(d2, r2)]$$

Bendro rezultato atpažinimas, kai vienas dalyvis:

Jei kiekvienas s priklauso sakinių aibei, dalyviai $d1$ priklauso dalyvių aibei, kiekvienas v priklauso varžybų aibei, rezultatai $r1$ priklauso rezultatų aibei, yra užfiksuota, kad sakiniui priklauso dalyvis $d1$, varžybos priklauso sakiniui, rezultatas $r1$ priklauso sakiniui, varžybas v laimėjo arba pralaimėjo dalyvis $d1$, tai dalyvio $d1$ rezultatas yra $r1$ (17).

$$\forall s \in S \forall d1 \in D \forall v \in V \forall r1 \in R [\text{priklauso}(s, d1) \wedge \text{priklauso}(v, s) \wedge \text{priklauso}(r1, s) \wedge [\text{laimėjo}(v, d1) \vee \text{pralaimėjo}(v, d1)]] \Rightarrow [\text{dalyvio_rezultatas}(d1, r1)] \quad (17)$$

Kėlinio rezultato atpažinimas kai du dalyviai:

Jei kiekvienas s priklauso sakinių aibei, dalyviai $d1$ ir $d2$ priklauso dalyvių aibei, kiekvienas k priklauso kėlinių aibei, rezultatai $kr1, kr2$ priklauso kėlinio rezultatų aibei, yra užfiksuota, kad sakiniui priklauso dalyviai $d1$ ir $d2$, varžybos priklauso sakiniui, rezultatai $kr1, kr2$ priklauso sakiniui, varžybas v laimėjo dalyvis $d1$ arba pralaimėjo dalyvis $d2$, tai dalyvio $d1$ kėlinio rezultatas yra $kr1$ ir dalyvio $d2$ kėlinio rezultatas yra $kr2$ (18).

$$\forall s \in S \forall d1, d2 \in D \forall v \in V \forall k \in K \forall kr1, kr2 \in KR [\text{priklauso}(s, d1) \wedge \text{priklauso}(s, d2) \wedge \text{priklauso}(v, s) \wedge \text{priklauso}(kr1, s) \wedge \text{priklauso}(kr2, s) \wedge [\text{laimėjo}(v, d1) \vee \text{pralaimėjo}(v, d2)]] \wedge \text{didesnis_už}(r1, r2)] \Rightarrow [\text{dalyvio_kėlinio_rezultatas}(d1, kr1, k) \wedge \text{dalyvio_kėlinio_rezultatas}(d2, kr2, k)] \quad (18)$$

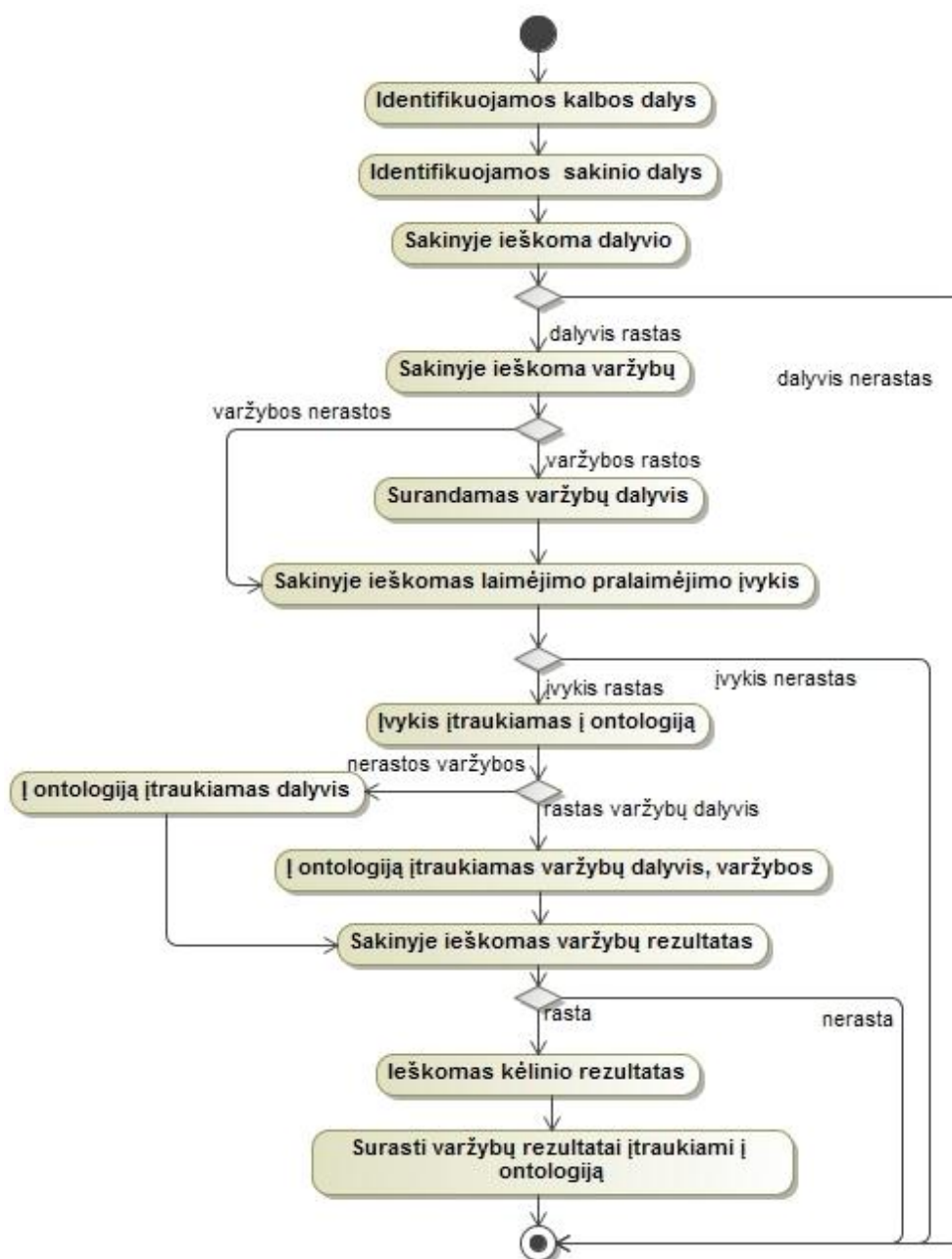
4.2. Įvykio atpažinimo metodikos taikymas

Prieš tai buvusiame skyriuje aprašytų taisyklių taikymo procesas pateikiamas žemiau esančiame 3.9 paveiksle.

Metodikos taikymo žingsniai:

1. Naudojant jau sukurtą morfologinio anotavimo įrankį Lemuoklis, identifikuojamos kalbos dalys.
2. Identifikuojamos sakinio dalys;
3. Naudojant taisyklių rinkinius „Komandos atpažinimas“ ir „Dalyvio atpažinimas“ surandamas sakinyje esantis dalyvis;
4. Jei yra rastas dalyvis, naudojant taisyklių rinkinį „Varžybų atpažinimas“, surandamos sakinyje paminėtos varžybos. Jei dalyvis nerastas, tolesni veiksmai nevykdomi;
5. Jei yra rasta dalyvis ir varžybos, naudojant taisyklių rinkinį „Varžybų dalyvio atpažinimas“, surandamas varžybų dalyvis.
6. Naudojant taisyklių rinkinį „Laimėjimo (pralaimėjimo) įvykio atpažinimas“, surandamas įvykis;

7. Jei įvykis randamas, jis įtraukiamas ontologiją, taip pat įtraukiami dalyvis ir varžybos ir susiejami per objektų savybes. Nurodomas laimėtojas ir pralaimėtojas.;
8. Jei įvykis nerastas, tolesni veiksmai nevykdomi;
9. Naudojant taisyklių rinkinį „Rezultato atpažinimas“, surandamas varžybų rezultatas.
10. Jeigu yra rastas varžybų rezultatas, naudojant taisyklių rinkinį „Kėlinio rezultato atpažinimas“, surandamas kėlinio rezultatas;
11. Jei yra rasta dalyvis, varžybos, varžybų ir (arba) kėlinio rezultatas, naudojant taisyklių rinkinį „Varžybų dalyvio rezultato atpažinimas“, surandamas varžybų dalyvio rezultatas. Į ontologiją įtraukiamas taškų gavimo įvykis, ji susiejamas su laimėjimo įvykiu, taškus surinkusia komanda.



4.8 pav. Įvykio atpažinimo metodikos taikymo procesas

5. EKSPERIMENTINIS TYRIMAS

5.1. Eksperimento planas

Prieš kuriant ontologiją buvo apibrėžti 11 klausimų į kuriuos turėtų atsakyti sporto įvykių ontologija. Eksperimentinio tyrimo metu šiems klausimams bus parašytos SPARQL užklauso. Įvykdžius užklauso, gauti rezultatai turėtų atsakyti į prieš tai užduotus klausimus. Tuo atveju bus galima daryti išvadą, kad sukurta sporto įvykių ontologija apima užsibrėžtą apimtį ir yra tinkama sporto įvykių semantiniam anotavimui.

Eksperimentas bus vykdomas vadovaujantis sukurta sporto įvykių atpažinimo tekste metodika.

5.2. Eksperimento vykdymas

Pirmiausia naujienų portaluose buvo surasti straipsniai, kurie bus naudojami eksperimente (8.1 priedas). Kadangi taisyklės skirtos atpažinti sporto įvykius sakinyje, todėl anotuojami tekstų fragmentai.

Toliau vykdomi veiksmai pagal sukurta įvykių atpažinimo tekste metodiką:

- Atliktas pasirinktų straipsnių morfologinis anotavimas. Morfologinio anotavimo rezultatai pateikiami 8.2 priede.
- Atliktas pasirinktų straipsnių sintaksinis anotavimas. Sintaksinio anotavimo rezultatai pateikiami 8.3 priede.
- Atliktas pasirinktų straipsnių sintaksinis anotavimas. Sintaksinio anotavimo rezultatai pateikiami 8.4 priede. Žemiau semantinio anotavimo procesas pirmam straipsniui (8.2 priedas):

- Pagal pirmąjį komandos atpažinimo šabloną atpažįstamos komandos:

Atpažintų esybių fragmentai	Tekstas su pažymėtais nagrinėjamais fragmentais
Type (Komanda, „Torpedo“) Type (Komanda, „CSKA“)	„Keturioliktą vietą Rusijos pirmo futbolo diviziono čempionate užimantis dviejų lietuvių atstovaujamas Maskvos „Torpedo“ klubas kontroliniame mače 2:1 (1:0) nugalėjo šalies aukščiausiosios lygos lyderius Maskvos CSKA futbolininkus.“[30].

- Pagal sporto renginio atpažinimo šablonus atpažįstamas sporto renginys:

Type (Macas, „kontroliniame mače“) Type (Rungtynės, „Rusijos pirmo futbolo diviziono čempionate“)	„Keturioliktą vietą Rusijos pirmo futbolo diviziono čempionate užimantis dviejų lietuvių atstovaujamas Maskvos „Torpedo“ klubas kontroliniame mače 2:1 (1:0) nugalėjo šalies aukščiausiosios lygos lyderius Maskvos CSKA futbolininkus.“ [30].
--	--

- Nustatomas varžybų dalyvis:

<p>varžybų_dalyvis („kontroliniame mače“, „Torpedo“)</p> <p>varžybų_dalyvis („kontroliniame mače“, „CSKA“)</p> <p>varžybų_dalyvis („Rusijos pirmo futbolo diviziono čempionate“, „Torpedo“)</p> <p>varžybų_dalyvis („Rusijos pirmo futbolo diviziono čempionate“, „CSKA“)</p> <p>dalyvavo_kur („Torpedo“, „Rusijos pirmo futbolo diviziono čempionate“)</p> <p>dalyvavo_kur („CSKA“, „Rusijos pirmo futbolo diviziono čempionate“)</p> <p>dalyvavo_kur („Torpedo“, „Rusijos pirmo futbolo diviziono čempionate“)</p> <p>dalyvavo_kur („CSKA“, „Rusijos pirmo futbolo diviziono čempionate“)</p>	<p>„Keturioliktą vietą Rusijos pirmo futbolo diviziono čempionate užimantis dviejų lietuvių atstovaujamas Maskvos „Torpedo“ klubas kontroliniame mače 2:1 (1:0) nugalėjo šalies aukščiausiosios lygos lyderius Maskvos CSKA futbolininkus.“ [30].</p>
---	---

- Naudojantis laimėjimų/ pralaimėjimų nustatymų taisyklėmis, nustatomas įvykis:

<p>Type (Įvykis, „nugalėjo“)</p> <p>nugalėjo_kas („kontroliniame mače“, „Torpedo“)</p> <p>pralaimėjo_kas („kontroliniame mače“, „CSKA“)</p>	<p>„Keturioliktą vietą Rusijos pirmo futbolo diviziono čempionate užimantis dviejų lietuvių atstovaujamas Maskvos „Torpedo“ klubas kontroliniame mače 2:1 (1:0) nugalėjo šalies aukščiausiosios lygos lyderius Maskvos CSKA futbolininkus.“ [30].</p>
---	---

- Naudojantis rezultatų atpažinimo taisyklėmis, atpažįstamas varžybų rezultatas:

<p>Rezultatas („kontroliniame mače“, „2“)</p> <p>Rezultatas („kontroliniame mače“, „1“)</p>	<p>„Keturioliktą vietą Rusijos pirmo futbolo diviziono čempionate užimantis dviejų lietuvių atstovaujamas Maskvos „Torpedo“ klubas kontroliniame mače 2:1 (1:0) nugalėjo šalies aukščiausiosios lygos lyderius Maskvos CSKA futbolininkus.“ [30].</p>
---	---

- Naudojantis kėlinio rezultatų atpažinimo taisyklėmis, atpažįstamas varžybų kėlinio rezultatas:

<p>KelinioRezultatas („kontroliniame mače“, „Pirmas“, „1“)</p> <p>KelinioRezultatas („kontroliniame mače“, „Pirmas“, „0“)</p>	<p>„Keturioliktą vietą Rusijos pirmo futbolo diviziono čempionate užimantis dviejų lietuvių atstovaujamas Maskvos „Torpedo“ klubas kontroliniame mače 2:1 (1:0) nugalėjo šalies aukščiausiosios lygos lyderius Maskvos CSKA futbolininkus.“ [30].</p>
---	---

- Naudojantis rezultatų atpažinimo taisyklėmis, atpažįstamas varžybų dalyvio rezultatas:

GautiTaskai („kontroliniame mače“, „Torpedo“, „2“)	„Keturioliką vietą Rusijos pirmo futbolo diviziono čempionate užimantis dviejų lietuvių
GautiTaskai („kontroliniame mače“, „CSKA“, „1“)	atstovaujamas Maskvos „Torpedo“ klubas kontroliniame mače 2:1 (1:0) nugalėjo šalies aukščiausiosios lygos lyderius Maskvos CSKA futbolininkus.“ [30].

Žemiau pateikti klausimai, į kuriuos turėjo atsakyti sukurta ontologija. Po kiekvienu klausimu pateikiama jam sudaryta SPARQL užklausa ir, įvykdžius tą užklausa, naudojant Protege įrankį, gauti rezultatai (5.1 – 5.11 pav.). Taip pat pateikiami paaiškinimai, kokius rezultatus turėtų pateikti užklausa. Pirmieji 4 klausimai pateikia atsakymus iš prieš tai tai semantiškai anotuoto 1 teksto, 5-6 klausimai pateikia atsakymus iš 2 teksto, 7-8 klausimai pateikia atsakymus iš 3 teksto, 9 pateikia atsakymus iš 4 teksto, 10 pateikia atsakymus iš 5 teksto, 11 pateikia atsakymus iš 6 teksto (8.2 priedas).

1. Kas laimėjo rungtynes tarp komandos A ir komandos B?

Turi būti surasti visi įvykiai, kuriuose užfiksuota pergalė ir dalyvavo komandos A ir B.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX iv: <http://www.semanticweb.org/ontologies/2013/2/OntologijaSporto1.owl#>
SELECT *
  WHERE {
    ?Pergalės_Įvykis rdf:type iv:Pergale.
    ?Rungtynių_pavadinimas iv:dalyvavo_kas iv:Torpedo.
    ?Rungtynių_pavadinimas iv:dalyvavo_kas iv:CSKA.
    ?Pergalės_Įvykis iv:vyko_įvykyje_kur ?Rungtynių_pavadinimas.
    ?Pergalės_Įvykis iv:nugalejo_kas ?kas_nugalėjo.
  }

```

Pergalės_įvykis	Rungtynių_pavadinimas	kas_nugalėjo
nugalėjo 2	mačas 2	CSKA
nugalėjo	kontroliniame mače	Torpedo

5.1 pav. Užklauskos „Kas laimėjo rungtynes tarp komandos A ir komandos B?“ rezultatas

2. Kokios komandos kovojo mače?

Turi būti surastos visos komandos, kurios dalyvavo sporto renginyje, kuris pažymėtas mačo klase.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX iv: <http://www.semanticweb.org/ontologies/2013/2/OntologijaSporto1.owl#>
SELECT *
  WHERE {
    ?Rungtynių_pavadinimas rdf:type iv:Macas.
    ?Rungtynių_pavadinimas iv:dalyvavo_kas ?Komanda.
    ?Komanda rdf:type iv:Komanda.
  }

```

}

Rungtynių pavadinimas	Komanda
kontroliniame mače	Torpedo
kontroliniame mače	CSKA
mačas 2	Torpedo
mačas 2	CSKA

5.2 pav. Užklauso „Kokios komandos kovojo mače?“ rezultatas

3. Kokios komandos kovojo mače X?

Turi būti surastos visos komandos, kurios dalyvavo nurodytame mače.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX iv: <http://www.semanticweb.org/ontologies/2013/2/OntologijaSporto1.owl#>
SELECT *
  WHERE {
    iv:mačas_2_ iv:dalyvavo_kas ? Komanda.
              ? Komanda rdf:type iv:Komanda.
  }

```

Komanda
Torpedo
CSKA

5.3 pav. Užklauso „Kokios komandos kovojo mače X?“ rezultatas

4. Koku rezultatu laimėjo komanda A prieš komandą B?

Turi būti surasti visi įvykiai, kuriuose užfiksuota komandos A pergalė prieš komandą B ir užfiksuotas rezultatas.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX iv: <http://www.semanticweb.org/ontologies/2013/2/OntologijaSporto1.owl#>
SELECT *
  WHERE {
    ?ivykis_pergales rdf:type iv:Pergale.
    ?ivykis_renginio iv:dalyvavo_kas iv:Torpedo.
    ?ivykis_renginio iv:dalyvavo_kas iv:CSKA.
    ?ivykis_pergales iv:vyko_įvykyje_kur ?ivykis_renginio.
    ?ivykis_pergales iv:nugalejo_kas iv:CSKA.
    ?ivykis_tasku_gavimo iv:vyko_įvykyje_kur ?ivykis_renginio.
    ?ivykis_tasku_gavimo rdf:type iv:GautiTaškai.
    ?ivykis_tasku_gavimo iv:taškai ?t1.
    ?ivykis_tasku_gavimo iv:gavo_taškus_kas ?komanda.
  }

```

ivykis_pergales	ivykis_renginio	ivykis_tasku_gavimo	t1	komanda
nugalėjo 2	mačas 2	gauti mačo 2 taškai	"1"@	Torpedo
nugalėjo 2	mačas 2	gavo taškus mače 2	"2"@	CSKA

5.4 pav. Užklauso „Koku rezultatu laimėjo komanda A prieš komandą B?“ rezultatas

5. Koks teisėjas teisėjavo rungtynėse X?

Turi būti surastas teisėjas, kuris teisėjavo nurodytoms varžyboms. Semantiškai anotuoto teksto pavyzdys pateiktas priede skyriuje


```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX iv: <http://www.semanticweb.org/ontologies/2013/2/OntologijaSporto1.owl#>
SELECT *
  WHERE {
    ?Ivykis rdf:type iv:Rungtynės.
    ?Ivykis iv:data ?Data.
  }

```

Ivykis	Data
NKL atkrintamųjų rungtynių	"2012-03-30"@

5.5 pav. Užklauso „Kada įvyko įvykis?“ rezultatas

6. Kas laimėjo pasaulio futbolo čempionatą ?

Turi būti surasta komanda, kuri nugalėjo pasaulio futbolo čempionatą. Semantiškai anotuoto teksto pavyzdys pateiktas priede skyriuje *Error! Reference source not found.*

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX iv: <http://www.semanticweb.org/ontologies/2013/2/OntologijaSporto1.owl#>
SELECT *
  WHERE {
    ?Ivykis_pergalės rdf:type iv:Pergale.
    ?Ivykis_pergalės iv:vyko_įvykyje_kur
iv:Pasaulio_futbolo_čempionato.
    ?Ivykis_pergalės iv:data ?Data.
    ?Ivykis_pergalės iv:nugalejo_kas ?Rinktinė.
  }

```

Ivykis_pergalės	Data	Rinktinė
nugalėjo pasaulio čempionate	"2010"@	Ispanijos futbolo rinktinė

5.6 pav. Užklauso „Kas laimėjo pasaulio futbolo čempionatą?“ rezultatas

7. Kas įmušė įvartį per varžybas X?

Turi būti surasti žaidėjai įmušę įvartį per nurodytas varžybas. Semantiškai anotuoto teksto pavyzdys pateiktas priede skyriuje *Error! Reference source not found.*

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX iv: <http://www.semanticweb.org/ontologies/2013/2/OntologijaSporto1.owl#>
SELECT *
  WHERE {
    ?Ivykis_taškų_gavimo iv:vyko_įvykyje_kur
iv:Pasaulio_futbolo_čempionato.
    ?Ivykis_taškų_gavimo iv:gavo_taškus_kas ?Žaidėjas.
  }

```

Ivykis_taškų_gavimo	Žaidėjas
pergalingą įvartį cempionate įsmeigė A.Iniesta	A. Iniesta

5.7 pav. Užklauso „Kas įmušė įvartį per varžybas X?“ rezultatas

8. Kiek žaidėjas Y įmetė taškų per rungtynes X?

Turi surasti kiek taškų surinko nurodytas žaidėjas per nurodytas rungtynes.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX iv: <http://www.semanticweb.org/ontologies/2013/2/OntologijaSporto1.owl#>
SELECT DISTINCT *
  WHERE {
    ?Ivykis_tasku_gavimo iv:vyko_ivykyje_kur iv:27-
ose_Turkijos_vyru_krepšinio_pirmenybiu_rungtynese.
    ?Ivykis_tasku_gavimo iv:gavo_taskus_kas ?Zaidėjas.
    ?Ivykis_tasku_gavimo rdf:type iv:GautiTaškai.
    ?Ivykis_tasku_gavimo iv:taškai ?Taškai.
  }

```

Ivykis_tasku_gavimo	Zaidėjas	Taškai
gavo taškus 27 turkijos rungtyne	Mindaugo Lukauskio	"11"@

5.8 pav. Užklauso „Kiek žaidėjas Y įmetė taškų per rungtynes X?“ rezultatas

9. Kokį rekordą pasiekė sportininkė Y per čempionatą X?

Turi surasti, kokį rekordą pasiekė nurodyta sportininkė per nurodytą čempionatą.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX iv: <http://www.semanticweb.org/ontologies/2013/2/OntologijaSporto1.owl#>
SELECT DISTINCT *
  WHERE {
    ?Rungtis iv:vyko_ivykyje_kur iv:XI_pasaulio_plaukimo_cempionate.
    ?Koks_Rekordas iv:pasiekė_kas iv:Rūta_Meilutytė.
    ?Koks_Rekordas iv:vyko_ivykyje_kur ?Rungtis.
    ?Koks_Rekordas iv:pasiektasRekordas ?Rekordas.
  }

```

Rungtis	Koks_Rekordas	Rekordas
100 metrų plaukimo laisvu stiliumi	pagerino Lietuvos rekordą	"53,54"@

5.9 pav. Užklauso „Kokį rekordą pasiekė sportininkė Y per čempionatą X?“ rezultatas

10. Kokį greitį pasiekė atletas rungtynėse X?

Turi surasti nurodytose rungtynėse dalyvavusių atletų greitį.

```

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX iv: <http://www.semanticweb.org/ontologies/2013/2/OntologijaSporto1.owl#>
SELECT *
  WHERE {
    ?Rungtynės iv:vyko_ivykyje_kur
iv:penktojo_pasaulio_atoralio_cempionato_.
    ?Greičio_Ivykis iv:vyko_ivykyje_kur ?Rungtynės.
    ?Greičio_Ivykis iv:pasiektasGreitis ?PasiektasGreitis.
  }

```

Rungtynės	Greičio_Ivykis	PasiektasGreitis
penktojo pasaulio atoralio čempionato	'1. Sebastien Loeb (Prancūzija)'	"4 val. 35 min. 56,7 sek"@

5.10 pav. Užklauso „Kokį greitį pasiekė atletas rungtynėse X?“ rezultatas

5.3. Eksperimento rezultatų analizė

Įvykdžius užklausas, gauti rezultatai atitinka klausimų atsakymus pagal tekstą, todėl galima daryti išvadą, kad sukurta sporto įvykių ontologija apima užsibrėžtą apimtį ir yra tinkama sporto įvykių semantiniam anotavimui.

Pagal sukurta metodiką, galima semantiškai suanotuoti tekstą, kai jame yra minima pergalė arba pralaimėjimas. Tačiau dar reikėtų įvesti naujų taisyklių, kurios būtų skirtos aptikti teisėjavimo, rekordo ar kokio nors rezultato pasiekimo įvykius, kurie nebūtinai turi būti pergalės įvykio subįvykiai.

5.4. Sprendimo taikymo rekomendacijos

Anotuojant tekstą naudojant sukurta sporto įvykių ontologiją, sporto įvykius reikia skaidyti į mažesnius susijusius įvykius. Tokius atveju lengviau bus parašyti užklausas ir išgauti tikslesnį norimą rezultatą. Šios sporto įvykių ontologijos pagrindu galima kurti ontologijas kitų sričių įvykiams, pvz.: muzikos įvykiai, politikos įvykiai.

6. REZULTATŲ APIBENDRINIMAS IR IŠVADOS

1. Išanalizavus semantinio anotavimo įrankius, nuspręsta sukurti metodiką, skirtą sporto įvykių atpažinimui. Kadangi dalis platformų semantiniam anotavimui naudoja taisyklėmis paremtus metodus, todėl ir kuriama sporto įvykių atpažinimo metodika bus paremta taisyklėmis, pagal kurias atpažįstamos ontologijos esybės.
2. Išanalizavus semantinio anotavimo procesą, sukurta metodika automatiniam sportinių įvykių atpažinimui sakiniuose. Ši metodika buvo panaudota, eksperimento metu, semantiniam teksto anotavimui. Išaiškėjo, kad šią metodiką reikėtų papildyti naujomis taisyklėmis, skirtomis, ne tik laimėjimų/pralaimėjimų atpažinimui, bet ir kitokio tipo sporto įvykiams.
3. Išanalizavus jau egzistuojančias BBC sporto ir Bendrąją įvykių ontologijas, sukurta ontologija, skirta sporto įvykių anotavimui. Taip pat nustatyti pagrindiniai sporto įvykių kūrimo principai, kuriais galima vadovautis kuriant kitų sričių įvykių ontologijas;
4. Eksperimentinio tyrimo metu sudarytos SPARQL užklauskos, skirtos įvykių paieškai ontologijoje. Atlikus paiešką, paaiškėjo, kad ontologija atsako į anksčiau užduotus klausimus, todėl galima spręsti, kad ontologija yra tinkama sporto įvykių semantiniam anotavimui.

7. LITERATŪRA

- [1] Andrew Sukhonogov, Serge Yablonsky, „Russian WordNet. From UML-notation to Database Implementation“. [Tinkle]. Prieiga per internetą: http://www.cs.vassar.edu/~ide/events/NLPXML03_review/papers/yablonsky.pdf. [Kreiptasi 2011-10-01];
- [2] Erika Rimkutė, „Dabartinės lietuvių kalbos gramatinių formų vartoseną morfologiškai anotuotame tekстыne“, 2006m. [Tinkle]. Prieiga per internetą: http://www.ebiblioteka.lt/resursai/LMA/Lituanistica/Lit_034_055.pdf. [Kreiptasi 2011-11-02].
- [3] Paul Buitelaar, „Natural Language Processing for the Semantic Web“, [Tinkle]. Prieiga per internetą: <http://www.deri.ie/fileadmin/documents/teaching/tutorials/DERI-Tutorial-NLP.final.pdf>. [Kreiptasi 2011-10-16];
- [4] Natalya F. Noy and Deborah L. McGuinness. „Ontology Development 101: A Guide to Creating Your First Ontology“, [Tinkle]. Prieiga per internetą: http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html. [Kreiptasi 2011-10-16]
- [5] „WordNet“, [Tinkle]. Prieiga per internetą: <http://wordnet.princeton.edu/>. [Kreiptasi 2011-11-01]
- [6] Eyal Oren, Knud Hinnerk Møller, Simon Scerri, Siegfried Handschuh, Michael Sintek. „What are Semantic Annotations?“. 2007m;
- [7] „Semantic Annotation“. [Tinkle]. Prieiga per internetą: <http://www.ontotext.com/kim/semantic-annotation>. [Kreiptasi 2011-11-01];
- [8] „KMI Semantic Annotation platform“. [Tinkle]. Prieiga per internetą: <http://annotation.semanticweb.org/Members/cobu/annotationtool.2004-10-04.2595486734>. [Kreiptasi 2011-11-01];
- [9] „MnM Semantic Annotation platform“. [Tinkle]. Prieiga per internetą: <http://annotation.semanticweb.org/Members/cobu/annotationtool.2004-10-08.8079905412>. [Kreiptasi 2011-11-01]
- [10] „GATE.“ [Tinkle]. Prieiga per internetą: <http://annotation.semanticweb.org/Members/lago/AnnotationTool.2003-08-25.1644>. [Kreiptasi 2011-11-01];
- [11] „OntoMat Annotizer“, [Tinkle]. Prieiga per internetą: <http://annotation.semanticweb.org/Members/cobu/AnnotationTool.2004-07-28.1138>. [Kreiptasi 2011-11-01];
- [12] Lawrence Reeve, Hyoil Han. Survey of Semantic Annotation Platforms. 2005m. Prieiga per psu.edu. [Kreiptasi 2011-11-07];
- [13] Lawrence Reeve, Hyoil Han. Technical Report: Semantic Annotation Platforms 2004 m. Prieiga per psu.edu. [Kreiptasi 2011-11-07];
- [14] David Taniar, Johanna Wenny Rahayu. „Web Semantic and Ontology.“ 2006 m.
- [15] Maria Vargas-Vera, Enrico Motta, John Domingue, Mattia Lanzoni, Arthur Stutt, Fabio Ciravegna. „MnM: Ontology Driven Semi-Automatic and Automatic Support for Semantic Markup“. 2002m.
- [16] GATE. Prieiga per internetą: <http://gate.ac.uk>. [Kreiptasi 2012-01-19];
- [17] Natalya F. Noy and Deborah L. McGuinness: Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, 2001.
- [18] „Sport Ontology“, 2011-02-17, [Tinkle]. Prieiga per internetą: <http://www.bbc.co.uk/ontologies/sport/2011-02-17.shtml>. [Kreiptasi 2012-11-03];
- [19] „Event Ontology“ 2007-10-25, [Tinkle]. Prieiga per internetą: <http://motools.sourceforge.net/event/event.html>. [Kreiptasi 2012-11-03];
- [20] „Protégé“, [Tinkle]. Prieiga per internetą: <http://protege.stanford.edu/overview/index.html>. [Kreiptasi 2013-04-04];

- [21] „TopBraid Composer“, [Tinkle]. Prieiga per internetą: http://www.topquadrant.com/products/TB_Composer.html. [Kreiptasi 2013-02-04];
- [22] „Russian National Corpus“, [Tinkle]. Prieiga per internetą: <http://www.ruscorpora.ru/en/index.html>. [Kreiptasi 2013-02-10];
- [23] „Corpus of contemporary American English“ [Tinkle]. Prieiga per internetą: <http://corpus.byu.edu/>. [Kreiptasi 2013-02-10];
- [24] Jem Rayfield, „Sports Refresh: Dynamic Semantic Publishing“, 2012-04-17. [Tinkle]. Prieiga per internetą: http://www.bbc.co.uk/blogs/bbcinternet/2012/04/sports_dynamic_semantic.html. [Kreiptasi 2013-02-10];
- [25] „Named Entity Tutorial“, [Tinkle]. Prieiga per internetą: <http://alias-i.com/lingpipe/demos/tutorial/ne/read-me.html>. [Kreiptasi 2013-03-24];
- [26] „Stanford NER“, [Tinkle]. Prieiga per internetą: <http://nlp.stanford.edu/software/CRF-NER.shtml>. [Kreiptasi 2013-03-24];
- [27] ANNIE, [Tinkle]. Prieiga per internetą: <http://www.aktors.org/technologies/annie/>. [Kreiptasi 2013-03-25];
- [28] „Brat features“, [Tinkle]. Prieiga per internetą: <http://brat.nlplab.org/>. [Kreiptasi 2013-03-25];
- [29] Eric Atwell, George Demetriou, John Hughes, Amanda Schiffrin, Clive Souter and Sean Wilcock: A comparative evaluation of modern English corpus grammatical annotation schemes. Centre for Computer Analysis of Language and Speech (CCALAS);
- [30] BNS, „Maskvos „Torpedo“ klubas kontroliniame mače įveikė Rusijos futbolo elito lyderį“, 2013-03-26. [Tinkle]. Prieiga per internetą: <http://sportas.delfi.lt/futbolas/maskvos-torpedo-klubas-kontroliniame-mace-iveike-rusijos-futbolo-elito-lyderi.d?id=60999617>. [Kreiptasi 2013-04-11];
- [31] „FRK Atletas Kaunas“, [Tinkle]. Prieiga per internetą: http://wiki.futbolinis.lt/index.php/FRK_Atletas_Kaunas. [Kreiptasi 2013-04-11];
- [32] Krašto naujienos, „Vyrų krepšinio pirmenybės. UAB „Aleksiškio“ krepšinio komandai aukso medaliai“, 2013-04-02. [Tinkle]. Prieiga per internetą: <http://xn--kratonaujienos-s0c.lt/2013/04/startavo-vyru-krepsinio-pirmenybes/>. [Kreiptasi 2013-04-11];
- [33] ELTA, „Tenisininkas Lukas Mugevičius baigė pasirodymą prestižinio „Orange Bowl“ turnyro vienetų varžybose“, 2012-12-06. [Tinkle]. Prieiga per internetą: <http://sportin.lt/naujienos/tenisas/tenisininkas-lukas-mugevicius-baige-pasirodyma-prestizinio-orange-bowl-turnyro-vienetu-varzybose/>. [žiūrėta 2013-04-11];
- [34] Alfa.lt „Džokovičius nutraukė įspūdingą ispano pergalių seriją Monte Karlo vyrų teniso turnyre“, [Tinkle]. Prieiga per internetą: <http://m.alfa.lt/naujienos/Sportas/16053/>. [Kreiptasi 2013-04-11];
- [35] 15min.lt, „Pergalių serija tęsiasi: Novakas Džokovičius iškopė į finalą Dubajuje“, 2013-03-01. [Tinkle]. Prieiga per internetą: <http://www.15min.lt/deuce/naujiena/profesionalus-sportas/naujienos/pergaliu-serija-tesiasi-novakas-dzokovicius-iskope-i-finala-dubajuje-338-312194>. [Kreiptasi 2013-04-11];
- [36] LFF, „LSFA "Select" taurė iškelia į Kazlų Rūdą“, 2013-03-11. [Tinkle]. Prieiga per internetą: <http://www.lff.lt/lt/article/lsfa-select-taure-iskeliavo-i-kazlu-ruda>. [Kreiptasi 2013-04-11];
- [37] Deldi.lt, „„Žalgiris“ Eurolygos sezoną baigė pergale prieš Vokietijos čempionus“, 2013-04-04. [Tinkle]. Prieiga per internetą: <http://sportas.delfi.lt/krepsinis/zalgiris-eurolygos-sezona-baige-pergale-pries-vokietijos-cempionus.d?id=61068787>. [Kreiptasi 2013-04-11];
- [38] BNS, „G.Paugaitės atstovaujama komanda nepateko į Prancūzijos krepšinio taurės turnyro finalą“, 2013-03-28. [Tinkle]. Prieiga per internetą: <http://sportas.delfi.lt/krepsinis/gpaugaites-atstovaujama-komanda-nepateko-i-prancuzijos-krepsinio-taures-turnyro-finala.d?id=61016865>. [Kreiptasi 2013-04-15];
- [39] BNS, „Vytenio Čižausko ekipa pralaimėjo ir Ispanijos krepšinio lygoje žengia šešiolikta“, 2013-04-20. [Tinkle]. Prieiga per internetą: <http://www.15min.lt/naujiena/sportas/krepsinis/vytenio-cizausko-ekipa-pralaimėjo-ir-ispnijos-krepsinio-lygoje-zengia-sesiolikta-23-327921>. [Kreiptasi 2013-05-10];
- [40] ELTA, „Tenisininkas Ričardas Berankis Indian Velse įveikė pirmąjį atrankos barjerą“, 2013-03-05. [Tinkle]. Prieiga per internetą: <http://sportin.lt/naujienos/tenisas/tenisininkas-ricardas-berankis-indian-velse-iveike-pirmaji-atrankos-barjera/>. [Kreiptasi 2013-05-10];

- [41] „D.Lobačevskė buvo nepralenkiama Hamburgo maratone!“, [Tinkle]. Prieiga per internetą: <http://m.sportas.lt/straipsnis/d-lobacevske-buvo-nepralenkiama-hamburgo-maratone-25537>. [Kreiptasi 2013-05-11];
- [42] Open Knowledge Base Connectivity 2.0.3, 1998. [Tinkle]. Prieiga per internetą: <http://www.ai.sri.com/~okbc/spec/okbc2/okbc2.html#SECTION00200000000000000000>. [Kreiptasi 2013-05-11];
- [43] Krepšinis.net „NKL rungtynėms teisėjavęs R.Brazauskas: likau maloniai nustebintas jaunų kolegų darbu“, 2012-03-30. [Tinkle]. Prieiga per internetą: <http://www.krepšinis.net/lietuva/nkl-rungtynems-teisejaves-r-brazauskas-likau-maloniai-nustebintas-jaunu-kolegu-darbu/196900>. [Kreiptasi 2013-05-19];
- [44] Matas Miknevičius, „Pasaulio futbolo čempionato finale – istorinė Ispanijos pergalė: Nyderlandai – Ispanija 0:1“, 2010-07-12. [Tinkle]. Prieiga per internetą: <http://www.lrytas.lt/-/12788697291278780209-pasaulio-futbolo-%C4%8Dempionato-finalas-nyderlandai-ispnija.htm>. [Kreiptasi 2013-05-18];
- [45] Erika Rimkutė, „Lietuvių kalbos tekstyno morfologinės analizės automatizavimas“, 2001-04-26. [Tinkle]. Prieiga per internetą: http://donelaitis.vdu.lt/publikacijos/sm_rimkute.pdf. [Kreiptasi 2013-05-18];
- [46] „Kompiuterinės lingvistikos centras“. [Tinkle]. Prieiga per internetą: <http://tekstynas.vdu.lt/>. [Kreiptasi 2013-05-18];
- [47] Michael Gruninger and Mark S, „Fox, Methodology for the Design and Evaluation of Ontologies“, 1995. [Tinkle]. Prieiga per internetą: <http://www.eil.utoronto.ca/enterprise-modelling/entmethod/gruninger-ijcai95.pdf>. [Kreiptasi 2013-05-18];
- [48] Mike Uschold and Martin King "Towards a Methodology for Building Ontologies", pristatyta: Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI-95, 1995;
- [49] BNS, „Turkijos krepšinio lygos mače M.Lukauskis įmetė vienuolika taškų“, 2013-04-28. [Tinkle]. Prieiga per internetą: <http://sportas.delfi.lt/krepšinis/turkijos-krepšinio-lygos-mace-mlukauskis-imate-vienuolika-tasku.d?id=61260327>. [Kreiptasi 2013-05-18];
- [50] Lrytas.lt, „Antroji R. Meilutytės diena pasaulio čempionate - pasiekti du pusfinaliai, bet viename iš jų nedalyvaus“, 2012-12-13. [Tinkle]. Prieiga per internetą: <http://www.lrytas.lt/sportas/startai/antroji-r-meilutytes-diena-pasaulio-cempionate-pasiekti-du-pusfinaliai-bet-viename-is-ju-nedalyvaus.htm>. [Kreiptasi 2013-05-18].
- [51] balsas.lt, „Devyniskart pasaulio čempionas S. Loebas aštuntą kartą iš eilės laimėjo Argentinos rali“, 2013-05-05. [Tinkle]. Prieiga per internetą: <http://www.balsas.lt/naujiena/731567/devyniskart-pasaulio-cempionas-s-loebas-astunta-karta-is-eiles-laimejo-argentin-os-rali> [Kreiptasi 2013-05-18].

8. PRIEDAI

8.1. priedas. Tekstų fragmentai

8.1 lentelė. Eksperimento tekstų fragmentai

1 tekstas: Keturiolikta vietą Rusijos pirmo futbolo diviziono čempionate užimantis dviejų lietuvių atstovaujamas Maskvos „Torpedo“ klubas kontroliniame mače 2:1 (1:0) nugalėjo šalies aukščiausiosios lygos lyderius Maskvos CSKA futbolininkus. [30]
2 tekstas: Ketvirtadienį Marijampolės žaidimų ir sporto centre susirinkę žiūrovai stebėjo ne tik „Sūduvos“ pergalę NKL atkrintamųjų rungtynių ketvirtfinalyje prieš Kauno „Žalgirį – Sabonio krepšinio centrą“, bet ir nepriekaištingą teisėjavimą: rungtynėms teisėjavo geriausiu Lietuvos krepšinio teisėju tituluojamas Romualdas Brazauskas. [43]
3 tekstas: Pasaulio futbolo čempionato auksą pirmą kartą šalies istorijoje iškovojo Ispanijos futbolo rinktinė. Pirmąjį kartą pasaulio futbolo čempionato finale žaidusi Ispanijos vienuolikė tik per antrąjį pratęsimo kėlinį palaužė Nyderlandų rinktinę ir iškovojo ilgai lauktą istorinę pergalę. Pergalingą įvartį smeigė patyręs ispanų žaidėjas A.Iniesta.[44]
4 tekstas: Šešiolikta pralaimėjimą 27-ose Turkijos vyrų krepšinio pirmenybių rungtynėse patyrė vienuolikta vietą užimantis Mindaugo Lukauskio Bursos „Tofaş“ klubas, šeštadienį namuose 73:75 (15:16, 18:24, 17:18, 23:17) turėjęs pripažinti keturioliktoje pozicijoje esančių Edirnės „Olin“ (8 pergalės ir 20 pralaimėjimų) krepšininkų pranašumą. Lietuvis aikštelės šeimininkams pelnė 11 taškų, po krepšiais atkovojo 6 ir perėmė 1 kamuolį, atliko 5 rezultatyvius perdavimus, po sykį suklydo bei prasižengė. [49]
5 tekstas: Ketvirtadienį Turkijoje, Stambule, XI pasaulio plaukimo čempionate trumpame (25 metrų) baseine Londono olimpinė čempionė 15-metė Rūta Meilutytė 100 metrų plaukimo laisvu stiliumi atrankos plaukime nuotolį įveikė per 53,54 sek., užėmė trečią vietą ir pateko į pusfinalį. Plaukikė pagerino Lietuvos rekordą, kuris priklausė Ditai Želvienei (54,30). [51]
6 tekstas: Šeštadienį pasibaigusio Argentinos ralio - penktojo pasaulio autoraliao čempionato (WRC) etapo - nugalėtoju aštuntą kartą iš eilės tapo nugalėtojo titulą ginantis devyniskart planetos čempionas prancūzas Sebastienas Loebas su "Citroen'u". Laimėjęs vos du greičio ruožus S. Loebas 55,0 sekundėmis aplenkė antrąją vietą užėmusį šio sezono lyderį bendrojoje įskaitoje tėvynainį Sebastieną Ogierą ("Volkswagen Polo"). Po keturių etapų S. Ogieras jau surinko 122 taškus ir 54 taškais lenkia į antrąją vietą pakilusį S. Loebą, kuris šiais metais dalyvaus tik keliuose planetos pirmenybių etapuose. Trečias Argentinoje šiemet liko suomis Jaris-Mattis Latvala ("Volkswagen Polo"), kuris nuo nugalėtojo atsiliko daugiau, kaip 2 minutėmis: 1. Sebastien Loeb (Prancūzija), "Citroen DS3" 4 val. 35 min. 56,7 sek. [52]

8.2. priedas. Tekstų morfologinio anotavimo rezultatas

8.2 lentelė. Morfologiškai anotuoti tekstai

1 tekstas: <word="Keturiolikta" lemma="keturioliktas" type="sktv., kelint., neįvardž., mot. g., vns., G."/> <space/> <word="vietą" lemma="vieta" type="dkt., mot. g., vns., G."/> <space/>

<word="Rusijos" lemma="Rusija" type="tikr. dkt., mot. g., vns., K."/>
 <space/>
 <word="pirmo" lemma="pirmas" type="sktv., kelint., nelygin. l., neįvardž., vyr. g., vns., K."/>
 <space/>
 <word="futbolo" lemma="futbolas" type="dkt., vyr. g., vns., K."/>
 <space/>
 <word="diviziono" lemma="divizionas" type="dkt., vyr. g., vns., K."/>
 <space/>
 <word="čempionate" lemma="čempionatas" type="dkt., vyr. g., vns., Vt."/>
 <space/>
 <word="užimantis" lemma="užimti(-ima,-ėmė)" type="dlv., teig., nesngr., veik. r, es. l., neįvardž., vyr. g., vns., V."/>
 <space/>
 <word="dviejų" lemma="du" type="sktv., kiek., vyr. g., K."/>
 <space/>
 <word="lietuvių" lemma="lietuvis" type="dkt., vyr. g., dgs., K."/>
 <space/>
 <word="atstovaujamas" lemma="atstovauti(-uja,-vo)" type="dlv., teig., nesngr., neveik. r, es. l., neįvardž., vyr. g., vns., V."/>
 <space/>
 <word="Maskvos" lemma="Maskva" type="tikr. dkt., mot. g., vns., K."/>
 <space/>
 <sep=","/>
 <word="Torpedo" lemma="Torpedo" type="nežinomas"/>
 <sep=""/>
 <space/>
 <word="klubas" lemma="klubas" type="dkt., vyr. g., vns., V."/>
 <space/>
 <word="kontroliniame" lemma="kontrolinis" type="bdv., teig, nelygin. l., neįvardž., vyr. g., vns., Vt."/>
 <space/>
 <word="mače" lemma="mačas" type="dkt., vyr. g., vns., Vt."/>
 <space/>
 <number="2"/>
 <sep=":"/>
 <number="1"/>
 <space/>
 <sep="("/>
 <number="1"/>
 <sep=":"/>
 <number="0"/>
 <sep=")"/>
 <space/>
 <word="nugalėjo" lemma="nugalėti(-i,-ėjo)" type="vksm., teig., nesngr., tiesiog. n., būt. k. l., vns., 3 asm."/>
 <space/>
 <word="šalis" lemma="šalis" type="dkt., mot. g., vns., K."/>
 <space/>
 <word="aukščiausiosios" lemma="aukštas" type="bdv., teig, aukšč. l., įvardž., mot. g., vns., K."/>
 <space/>
 <word="lygos" lemma="lyga" type="dkt., mot. g., vns., K."/>
 <space/>
 <word="lyderius" lemma="lyderis" type="dkt., vyr. g., dgs., G."/>
 <space/>
 <word="Maskvos" lemma="Maskva" type="tikr. dkt., mot. g., vns., K."/>
 <space/>
 <word="CSKA" lemma="CSKA" type="akronim."/>
 <space/>
 <word="futbolininkus" lemma="futbolininkas" type="dkt., vyr. g., dgs., G."/>
 <sep="."/>
 <sep=""/>
 <sep="["/>
 <p/>

2 tekstas:

<word="Ketvirtadienį" type="dkt., vyr. g., vns., G."/>
 <space/>

<word="Marijampolės" type="tikr. dkt., mot. g., vns., K."/>
 <space/>
 <word="žaidimų" type="dkt., vyr. g., dgs., K."/>
 <space/>
 <word="ir" type="jng."/>
 <space/>
 <word="sporto" type="dkt., vyr. g., vns., K."/>
 <space/>
 <word="centre" type="dkt., vyr. g., vns., Vt."/>
 <space/>
 <word="susirinkę" type="dlv., teig., sngr., veik. r, būt. k. l., neįvardž., vyr. g., dgs., V."/>
 <space/>
 <word="žiūrovai" type="dkt., vyr. g., dgs., V."/>
 <space/>
 <word="stebėjo" type="vksm., teig., nesngr., tiesiog. n., būt. k. l., dgs., 3 asm."/>
 <space/>
 <word="ne tik" type="jng."/>
 <space/>
 <sep="<"/>
 <word="Sūduvos" type="tikr. dkt., mot. g., vns., K."/>
 <sep=">"/>
 <space/>
 <word="pergalę" type="dkt., mot. g., vns., G."/>
 <space/>
 <word="NKL" type="nežinomas"/>
 <space/>
 <word="atkrinamųjų" type="dlv., teig., nesngr., neveik. r, es. l., įvardž., vyr. g., dgs., K."/>
 <space/>
 <word="rungtynių" type="dkt., mot. g., dgs., K."/>
 <space/>
 <word="ketvirtfinalyje" type="dkt., vyr. g., vns., Vt."/>
 <space/>
 <word="prieš" type="prl."/>
 <space/>
 <word="Kauno" type="tikr. dkt., vyr. g., vns., K."/>
 <space/>
 <sep="<"/>
 <word="Žalgirį" type="tikr. dkt., vyr. g., vns., G."/>
 <space/>
 <word="Sabonio" type="nežinomas"/>
 <space/>
 <word="krepšinio" type="dkt., vyr. g., vns., K."/>
 <space/>
 <word="centrą" type="dkt., vyr. g., vns., G."/>
 <sep=">"/>
 <sep=","/>
 <space/>
 <word="bet ir" type="jng."/>
 <space/>
 <word="nepriekaištingą" type="bdv., neig., nelygin. l., neįvardž., vyr. g., vns., G."/>
 <space/>
 <word="teisėjavimą" type="dkt., vyr. g., vns., G."/>
 <sep=":"/>
 <space/>
 <word="rungtynėms" type="dkt., mot. g., dgs., N."/>
 <space/>
 <word="teisėjavo" type="vksm., teig., nesngr., tiesiog. n., būt. k. l., vns., 3 asm."/>
 <space/>
 <word="geriausiai" type="bdv., teig. aukšč. l., neįvardž., vyr. g., vns., Įn."/>
 <space/>
 <word="Lietuvos" type="tikr. dkt., mot. g., vns., K."/>
 <space/>
 <word="krepšinio" type="dkt., vyr. g., vns., K."/>

<space/>
<word="teisėju" type="dkt., vyr. g., vns., Įn."/>
<space/>
<word="tituluojamas" type="dlv., teig., nesngr., neveik. r, es. l., nežvardž., vyr. g., vns., V."/>
<space/>
<word="Romualdas" type="tikr. dkt., vyr. g., vns., V."/>
<space/>
<word="Brazauskas" type="nežinomas"/>
<sep="."/>
<p/>

3 tekstas:

<word="Pasaulio" type="dkt., vyr. g., vns., K."/>
<space/>
<word="futbolo" type="dkt., vyr. g., vns., K."/>
<space/>
<word="čempionato" type="dkt., vyr. g., vns., K."/>
<space/>
<word="auksą" type="dkt., vyr. g., vns., G."/>
<space/>
<word="pirmą" type="sktv., kelint., nelygin. l., nežvardž., vyr. g., vns., G."/>
<space/>
<word="kartą" type="dkt., vyr. g., vns., G."/>
<space/>
<word="šalies" type="dkt., mot. g., vns., K."/>
<space/>
<word="istorijoje" type="dkt., mot. g., vns., Vt."/>
<space/>
<word="iškovojo" type="vksm., teig., nesngr., tiesiog. n., būt. k. l., vns., 3 asm."/>
<space/>
<word="Ispanijos" type="tikr. dkt., mot. g., vns., K."/>
<space/>
<word="futbolo" type="dkt., vyr. g., vns., K."/>
<space/>
<word="rinktinė" type="dkt., mot. g., vns., V."/>
<sep="."/>
<space/>
<word="Pirmąjį" type="sktv., kelint., nelygin. l., įvardž., vyr. g., vns., G."/>
<space/>
<word="kartą" type="dkt., vyr. g., vns., G."/>
<space/>
<word="pasaulio" type="dkt., vyr. g., vns., K."/>
<space/>
<word="futbolo" type="dkt., vyr. g., vns., K."/>
<space/>
<word="čempionato" type="dkt., vyr. g., vns., K."/>
<space/>
<word="finale" type="dkt., vyr. g., vns., Vt."/>
<space/>
<word="žaidusi" type="dlv., teig., nesngr., veik. r, būt. k. l., nežvardž., mot. g., vns., V."/>
<space/>
<word="Ispanijos" type="tikr. dkt., mot. g., vns., K."/>
<space/>
<word="vienuolikė" type="dkt., mot. g., vns., V."/>
<space/>
<word="tik" type="prv., teig., nelygin. l."/>
<space/>
<word="per" type="dll."/>
<space/>
<word="antrąjį" type="sktv., kelint., įvardž., vyr. g., vns., G."/>
<space/>
<word="pratęsimo" type="dkt., vyr. g., vns., K."/>
<space/>
<word="kėlinį" type="dkt., vyr. g., vns., G."/>

<space/>
 <word="palaužė" type="vksm., teig., nesngr., tiesiog. n., būt. k. l., vns., 3 asm."/>
 <space/>
 <word="Nyderlandų" type="nežinomas"/>
 <space/>
 <word="rinktinė" type="dkt., mot. g., vns., G."/>
 <space/>
 <word="ir" type="jng."/>
 <space/>
 <word="iškovojo" type="vksm., teig., nesngr., tiesiog. n., būt. k. l., vns., 3 asm."/>
 <space/>
 <word="ilgai" type="prv., teig., nelygin. l."/>
 <space/>
 <word="lauktą" type="bdv., teig, nelygin. l., neįvardž., vyr. g., vns., G."/>
 <space/>
 <word="istorinę" type="bdv., teig, nelygin. l., neįvardž., mot. g., vns., G."/>
 <space/>
 <word="pergalę" type="dkt., mot. g., vns., G."/>
 <sep="."/>
 <space/>
 <word="Pergalingą" type="bdv., teig, nelygin. l., neįvardž., vyr. g., vns., G."/>
 <space/>
 <word="įvartį" type="dkt., vyr. g., vns., G."/>
 <space/>
 <word="smeigė" type="dkt., mot. g., vns., V."/>
 <space/>
 <word="patyręs" type="dlv., teig., nesngr., veik. r, būt. k. l., neįvardž., vyr. g., vns., V."/>
 <space/>
 <word="ispanų" type="dkt., vyr. g., dgs., K."/>
 <space/>
 <word="žaidėjas" type="dkt., vyr. g., vns., V."/>
 <space/>
 <word="A" type="jst."/>
 <sep="."/>
 <word="Iniesta" type="nežinomas"/>
 <sep="."/>
 <p/>

4 tekstas:

<word="Šešioliktą" lemma="šešioliktas" type="sktv., kelint., neįvardž., vyr. g., vns., G."/>
 <space/>
 <word="pralaimėjimą" lemma="pralaimėjimas" type="dkt., vyr. g., vns., G."/>
 <space/>
 <number="27"/>
 <sep="."/>
 <word="ose" lemma="ose" type="nežinomas"/>
 <space/>
 <word="Turkijos" lemma="Turkija" type="tikr. dkt., mot. g., vns., K."/>
 <space/>
 <word="vyrų" lemma="vyras" type="dkt., vyr. g., dgs., K."/>
 <space/>
 <word="krepšinio" lemma="krepšinis" type="dkt., vyr. g., vns., K."/>
 <space/>
 <word="pirmenybių" lemma="pirmenybė" type="dkt., mot. g., dgs., K."/>
 <space/>
 <word="rungtynėse" lemma="rungtynės" type="dkt., mot. g., dgs., Vt."/>
 <space/>
 <word="patyrė" lemma="patirti(-iria,-yrė)" type="vksm., teig., nesngr., tiesiog. n., būt. k. l., dgs., 3 asm."/>
 <space/>
 <word="vieniuliktą" lemma="vieniuliktas" type="sktv., kelint., neįvardž., mot. g., vns., G."/>
 <space/>
 <word="vietą" lemma="vieta" type="dkt., mot. g., vns., G."/>
 <space/>
 <word="užimantis" lemma="užimti(-ima,-ėmė)" type="dlv., teig., nesngr., veik. r, es. l., neįvardž., vyr. g., vns., V."/>

```

<space/>
<word="Mindaugo" lemma="Mindaugas" type="tikr. dkt., vyr. g., vns., K."/>
<space/>
<word="Lukauskio" lemma="Lukauskio" type="nežinomas"/>
<space/>
<word="Bursos" lemma="bursa" type="dkt., mot. g., vns., K."/>
<space/>
<sep=","/>
<word="Tofaš" lemma="Tofaš" type="nežinomas"/>
<sep="<">
<space/>
<word="klubas" lemma="klubas" type="dkt., vyr. g., vns., V."/>
<sep=","/>
<space/>
<word="šeštadienį" lemma="šeštadienis" type="dkt., vyr. g., vns., G."/>
<space/>
<word="namuose" lemma="namai" type="dkt., vyr. g., dgs., Vt."/>
<space/>
<number="73"/>
<sep=":"/>
<number="75"/>
<space/>
<sep="("/>
<number="15"/>
<sep=":"/>
<number="16"/>
<sep=","/>
<space/>
<number="18"/>
<sep=":"/>
<number="24"/>
<sep=","/>
<space/>
<number="17"/>
<sep=":"/>
<number="18"/>
<sep=","/>
<space/>
<number="23"/>
<sep=":"/>
<number="17"/>
<sep=")"/>
<space/>
<word="turėjęs" lemma="turėti(-i,-ėjo)" type="dlv., teig., nesngr., veik. r, būt. k. l., neįvardž., vyr. g., vns., V."/>
<space/>
<word="pripažinti" lemma="pripažinti(-įsta,-ino)" type="bendr., teig., nesngr."/>
<space/>
<word="keturioliktoje" lemma="keturioliktas" type="sktv., kelint., neįvardž., mot. g., vns., Vt."/>
<space/>
<word="pozicijoje" lemma="pozicija" type="dkt., mot. g., vns., Vt."/>
<space/>
<word="esančių" lemma="būti(yra,buvo)" type="dlv., teig., nesngr., veik. r, es. l., neįvardž., vyr. g., dgs., K."/>
<space/>
<word="Edirnės" lemma="Edirnės" type="nežinomas"/>
<space/>
<sep=","/>
<word="Olin" lemma="Olin" type="nežinomas"/>
<sep="<">
<space/>
<sep="("/>
<number="8"/>
<space/>
<word="pergalės" lemma="pergalė" type="dkt., mot. g., vns., K."/>

```

<space/>
 <word="ir" lemma="ir" type="jng."/>
 <space/>
 <number="20"/>
 <space/>
 <word="pralaimėjimų" lemma="pralaimėjimas" type="dkt., vyr. g., dgs., K."/>
 <sep="")"/>
 <space/>
 <word="krepšininkų" lemma="krepšininkas" type="dkt., vyr. g., dgs., K."/>
 <space/>
 <word="pranašumą" lemma="pranašumas" type="dkt., vyr. g., vns., G."/>
 <sep=".""/>
 <space/>
 <word="Lietuvis" lemma="lietuvis" type="dkt., vyr. g., vns., V."/>
 <space/>
 <word="aikštelės" lemma="aikštelė" type="dkt., mot. g., vns., K."/>
 <space/>
 <word="šeimininkams" lemma="šeimininkas" type="dkt., vyr. g., dgs., N."/>
 <space/>
 <word="pelnė" lemma="pelnyti(-o,-ė)" type="vksm., teig., nesngr., tiesiog. n., būt. k. l., vns., 3 asm."/>
 <space/>
 <number="11"/>
 <space/>
 <word="taškų" lemma="taškas" type="dkt., vyr. g., dgs., K."/>
 <sep=",""/>
 <space/>
 <word="po" lemma="po" type="prl."/>
 <space/>
 <word="krepšiais" lemma="krepšys" type="dkt., vyr. g., dgs., Įn."/>
 <space/>
 <word="atkovojo" lemma="atkovoti(-ja,-jo)" type="vksm., teig., nesngr., tiesiog. n., būt. k. l., dgs., 3 asm."/>
 <space/>
 <number="6"/>
 <space/>
 <word="ir" lemma="ir" type="jng."/>
 <space/>
 <word="perėmė" lemma="perimti(-ima,-ėmė)" type="vksm., teig., nesngr., tiesiog. n., būt. k. l., vns., 3 asm."/>
 <space/>
 <number="1"/>
 <space/>
 <word="kamuolių" lemma="kamuolys" type="dkt., vyr. g., vns., G."/>
 <sep=",""/>
 <space/>
 <word="atliko" lemma="atlikti(-ieka,-iko)" type="vksm., teig., nesngr., tiesiog. n., būt. k. l., vns., 3 asm."/>
 <space/>
 <number="5"/>
 <space/>
 <word="rezultatyvius" lemma="rezultatyvus" type="bdv., teig., nelygin. l., neįvardž., vyr. g., dgs., G."/>
 <space/>
 <word="perdavimus" lemma="perdavimas" type="dkt., vyr. g., dgs., G."/>
 <sep=",""/>
 <space/>
 <word="po" lemma="po" type="prl."/>
 <space/>
 <word="sykių" lemma="sykis" type="dkt., vyr. g., vns., G."/>
 <space/>
 <word="suklydo" lemma="suklysti(-sta,-do)" type="vksm., teig., nesngr., tiesiog. n., būt. k. l., vns., 3 asm."/>
 <space/>
 <word="bei" lemma="bei" type="jng."/>
 <space/>
 <word="prasižengė" lemma="prasižengti(-gia,-gė)" type="vksm., teig., sngr., tiesiog. n., būt. k. l., vns., 3 asm."/>
 <sep=".""/>
 <p/>

8.3. priedas. Tekstų sintaksinio anotavimo rezultatas

Veiksny - geltonai paryškintas tekstas, tarinys – žaliai.

8.3 lentelė. Sintaksiškai anotuotas tekstas

<p>1 tekstas:</p> <p>Keturioliką vietą Rusijos pirmo futbolo diviziono čempionate užimantis dviejų lietuvių atstovaujamas Maskvos „Torpedo“ klubas kontroliniame mače 2:1 (1:0) nugalėjo šalies aukščiausios lygos lyderius Maskvos CSKA futbolininkus. [30]</p>
<p>2 tekstas:</p> <p>Ketvirtadienį Marijampolės žaidimų ir sporto centre susirinkę žiūrovai stebėjo ne tik “Sūduvos” pergalę NKL atkrintamųjų rungtynių ketvirtfinalyje prieš Kauno “Žalgirį – Sabonio krepšinio centrą”, bet ir nepriekaištingą teisėjavimą: rungtynėms teisėjo geriausiu Lietuvos krepšinio teisėju tituluojamas Romualdas Brazauskas. [43]</p>
<p>3 tekstas:</p> <p>Pasaulio futbolo čempionato auksą pirmą kartą šalies istorijoje iškovojo Ispanijos futbolo rinktinė. Pirmąjį kartą pasaulio futbolo čempionato finale žaidusi Ispanijos vienuolikė tik per antrąjį pratęsimo kėlinį palaužė Nyderlandų rinktinę ir iškovojo ilgai lauktą istorinę pergalę. Pergalingą įvartį smeigė patyręs ispanų žaidėjas A.Iniesta. [44]</p>
<p>4 tekstas:</p> <p>Šešioliką pralaimėjimą 27-ose Turkijos vyrų krepšinio pirmenybių rungtynėse patyrė vienuoliką vietą užimantis Mindaugo Lukauskio Bursos „Tofaş“ klubas, šeštadienį namuose 73:75 (15:16, 18:24, 17:18, 23:17) turėjęs pripažinti keturioliktoje pozicijoje esančių Edirnės „Olin“ (8 pergalės ir 20 pralaimėjimų) krepšininkų pranašumą. Lietuvis aikštelės šeimininkams pelnė 11 taškų, po krepšiais atkovojo 6 ir perėmė 1 kamuolį, atliko 5 rezultatyvius perdavimus, po syki suklydo bei prasižengė. [49]</p>

8.4. priedas. Tekstų semantinio anotavimo pavyzdžiai

Žemiau esančiose lentelėse pateikiami pavyzdys, kaip, naudojantis sukurta ontologija ir aprašytomis taisyklėmis, galima anotuoti tekstą.

8.4 lentelė. 2 tekstas, pažymėtos pagrindinės anotacijos.

Ontologijos fragmentas	Tekstas su pažymėtais nagrinėjama fragmentais
Type (Komanda, „Sūduvos“) Type (Komanda, „Žalgirį – Sabonio krepšinio centrą“)	„Ketvirtadienį Marijampolės žaidimų ir sporto centre susirinkę žiūrovai stebėjo ne tik “Sūduvos” pergalę NKL atkrintamųjų rungtynių ketvirtfinalyje prieš Kauno “Žalgirį – Sabonio krepšinio centrą”, bet ir nepriekaištingą teisėjavimą: rungtynėms teisėjo geriausiu Lietuvos krepšinio teisėju tituluojamas Romualdas Brazauskas.“ [43]
Type (Įvykis, „pergalė“) Teisėjo_kas („NKL atkrintamųjų rungtynių“, „Romualdas Brazauskas“)	„Ketvirtadienį Marijampolės žaidimų ir sporto centre susirinkę žiūrovai stebėjo ne tik “Sūduvos” pergalę NKL atkrintamųjų rungtynių ketvirtfinalyje prieš Kauno “Žalgirį – Sabonio krepšinio centrą”, bet ir nepriekaištingą teisėjavimą: rungtynėms teisėjo geriausiu Lietuvos krepšinio teisėju tituluojamas Romualdas Brazauskas. “

	[43]
Type (Rungtynes, „NKL atkrintamųjų rungtynių“)	„Ketvirtadienį Marijampolės žaidimų ir sporto centre susirinkę žiūrovai stebėjo ne tik “Sūduvos” pergalę NKL atkrintamųjų rungtynių ketvirtfinalyje prieš Kauno “Žalgirį – Sabonio krepšinio centrą”, bet ir nepriekaištingą teisėjavimą: rungtynėms teisėjavo geriausiu Lietuvos krepšinio teisėju tituluojamas Romualdas Brazauskas.“ [43]
Type (Teisejas, „Romualdas Brazauskas“)	„Ketvirtadienį Marijampolės žaidimų ir sporto centre susirinkę žiūrovai stebėjo ne tik “Sūduvos” pergalę NKL atkrintamųjų rungtynių ketvirtfinalyje prieš Kauno “Žalgirį – Sabonio krepšinio centrą”, bet ir nepriekaištingą teisėjavimą: rungtynėms teisėjavo geriausiu Lietuvos krepšinio teisėju tituluojamas Romualdas Brazauskas .“ [43]

8.5 lentelė. 3 tekstas, pažymėtos pagrindinės anotacijos.

Ontologija	Tekstas su pažymėtais nagrinėjamais fragmentais
Type (Rinktinė, „Ispanijos“) Type (Rinktinė, „Nyderlandų“)	„Pasaulio futbolo čempionato auksą pirmą kartą šalies istorijoje iškovojo Ispanijos futbolo rinktinė. Pirmąjį kartą pasaulio futbolo čempionato finale žaidusi Ispanijos vienuolikė tik per antrąjį pratęsimo kėlinį palaužė Nyderlandų rinktinę ir iškovojo ilgai lauktą istorinę pergalę. Pergalingą įvartį smeigė patyręs ispanų žaidėjas A.Iniesta.“ [44]
Type (Įvykis, „pergalė“)	„Pasaulio futbolo čempionato auksą pirmą kartą šalies istorijoje iškovojo Ispanijos futbolo rinktinė. Pirmąjį kartą pasaulio futbolo čempionato finale žaidusi Ispanijos vienuolikė tik per antrąjį pratęsimo kėlinį palaužė Nyderlandų rinktinę ir iškovojo ilgai lauktą istorinę pergalę . Pergalingą įvartį smeigė patyręs ispanų žaidėjas A.Iniesta.“ [44]
Type (Rungtynes, „pasaulio futbolo čempionato“)	„Pirmąjį kartą pasaulio futbolo čempionato finale žaidusi Ispanijos vienuolikė tik per antrąjį pratęsimo kėlinį palaužė Nyderlandų rinktinę ir iškovojo ilgai lauktą istorinę pergalę. Pergalingą įvartį smeigė patyręs ispanų žaidėjas A.Iniesta.“ [44]
Type (Žaidėjas, „A.Iniesta“)	„Pasaulio futbolo čempionato auksą pirmą kartą šalies istorijoje iškovojo Ispanijos futbolo rinktinė. Pirmąjį kartą pasaulio futbolo čempionato finale žaidusi Ispanijos vienuolikė tik per antrąjį pratęsimo kėlinį palaužė Nyderlandų rinktinę ir iškovojo ilgai lauktą istorinę pergalę. Pergalingą įvartį smeigė patyręs ispanų žaidėjas A.Iniesta .“ [44]
Type (GautiTaskai, „Pergalingą įvartį smeigė A.Iniesta“)	„Pasaulio futbolo čempionato auksą pirmą kartą šalies istorijoje iškovojo Ispanijos futbolo rinktinė. Pirmąjį kartą pasaulio

	futbolo čempionato finale žaidusi Ispanijos vienuolikė tik per antrąjį pratęsimo kėlinį palaužė Nyderlandų rinktinę ir iškovojo ilgai lauktą istorinę pergalę. Pergalingą įvartį smeigė patyręs ispanų žaidėjas A.Iniesta.“ [44]
--	--

8.6 lentelė. 4 tekstas, pažymėtos pagrindinės anotacijos.

Ontologija	Tekstas su pažymėtais nagrinėjamaais fragmentais
Type (Komanda, „Tofaş“) Type (Rinktinė, „Olin“)	Šešiolikta pralaimėjimą 27-ose Turkijos vyrų krepšinio pirmenybių rungtynėse patyrė vienuolikta vietą užimantis Mindaugo Lukauskio Bursos „Tofaş“ klubas, šeštadienį namuose 73:75 (15:16, 18:24, 17:18, 23:17) turėjęs pripažinti keturioliktoje pozicijoje esančių Edirnės „Olin“ (8 pergalės ir 20 pralaimėjimų) krepšininkų pranašumą. Lietuvis aikštelės šeimininkams pelnė 11 taškų, po krepšiais atkovojo 6 ir perėmė 1 kamuolį, atliko 5 rezultatyvius perdavimus, po syki suklydo bei prasižengė. [49]
Type (Rungtynes, „27-ose Turkijos vyrų krepšinio pirmenybių rungtynėse“)	Šešiolikta pralaimėjimą 27-ose Turkijos vyrų krepšinio pirmenybių rungtynėse patyrė vienuolikta vietą užimantis Mindaugo Lukauskio Bursos „Tofaş“ klubas, šeštadienį namuose 73:75 (15:16, 18:24, 17:18, 23:17) turėjęs pripažinti keturioliktoje pozicijoje esančių Edirnės „Olin“ (8 pergalės ir 20 pralaimėjimų) krepšininkų pranašumą. Lietuvis aikštelės šeimininkams pelnė 11 taškų, po krepšiais atkovojo 6 ir perėmė 1 kamuolį, atliko 5 rezultatyvius perdavimus, po syki suklydo bei prasižengė. [49]
Type (Gauti taškai, „Mindaugo Lukauskio“, „11“)	Šešiolikta pralaimėjimą 27-ose Turkijos vyrų krepšinio pirmenybių rungtynėse patyrė vienuolikta vietą užimantis Mindaugo Lukauskio Bursos „Tofaş“ klubas, šeštadienį namuose 73:75 (15:16, 18:24, 17:18, 23:17) turėjęs pripažinti keturioliktoje pozicijoje esančių Edirnės „Olin“ (8 pergalės ir 20 pralaimėjimų) krepšininkų pranašumą. Lietuvis aikštelės šeimininkams pelnė 11 taškų, po krepšiais atkovojo 6 ir perėmė 1 kamuolį, atliko 5 rezultatyvius perdavimus, po syki suklydo bei prasižengė. [49]