



Совместный выпуск

Научный журнал

Сентябрь, 2018 г.

№ 3. Том 1. Часть 3

ВЕСТНИК

Восточно-Казахстанского
государственного технического
университета им. Д. Серикбаева

ВЫЧИСЛИТЕЛЬНЫЕ ТЕХНОЛОГИИ

Институт вычислительных технологий
Сибирского отделения РАН

ИВТ





СибГУТИ

СИБИРСКИЙ
ГОСУДАРСТВЕННЫЙ
УНИВЕРСИТЕТ
ТЕЛЕКОММУНИКАЦИЙ
И ИНФОРМАТИКИ



**Novosibirsk State
Technical University**

HLRIS
High-Performance Computing Center | Stuttgart



N* Новосибирский
государственный
университет
***НАСТОЯЩАЯ НАУКА**



**University of Pristina
Kosovska Mitrovica**



МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ «Вычислительные и информационные технологии в науке, технике и образовании»



25-28 сентября

**Республика Казахстан, г. Усть-Каменогорск,
ул. Серикбаева, 19, 2018 г.**

ВЕСТНИК

**Восточно-Казахстанского государственного
технического университета
им. Д. Серикбаева**

Главный редактор
доктор биологических наук, профессор
Ж.К. Шаймарданов



ВЫЧИСЛИТЕЛЬНЫЕ ТЕХНОЛОГИИ

**Институт вычислительных технологий
Сибирского отделения РАН**

Главный редактор
академик
Ю.И. Шокин



по материалам Международной конференции
«Вычислительные и информационные технологии
в науке, технике и образовании»
(CITech-2018)
25-28 сентября 2018 года

Том I. Часть III

Усть-Каменогорск-Новосибирск, 2018

Статистикалық әдістер негізінде пәндік аймақ онтологиясын құру

Б.Б. Досанов, М.Е. Мансурова, А.Б. Нугуманова

Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан
С. Аманжолов атындағы Шығыс Қазақстан мемлекеттік университеті,
Өскемен, Қазақстан
dossanbekzhan@gmail.com

Аңдатпа Бұл мақалада белгілі бір пәндік аймақ үшін, формальді түсінік негізінде, Бутстраппинг технологиясы арқылы берілген сөйлем ішінен түсініктерді алу комбинациясының жаңа тәсілі жайында. Формальді түсінік теориясы автоматты түрде пәндік аймақ үшін түсініктерді алуға арналған мықты аппарат, бірақ кіріс дерек негізінде пәндік аймақ деректер қоры толық, нақты әрі қуыс болмауы керек. Берілген сөйлем ішінен пәндік аймақ негізінде мұндай толық деректер қорын алу өте қиын, себебі сөйлем ішінен кейбір түсініктер үшін сиректілік байқалады, яғни бастапқы деректер қорында бос қуыстық байқалады. Осыған сәйкес, Вебтегі фрагменттелген ақпараттарды интеллектуалды іздестіруді қамтамасыз ете отырып, Бутстраппинг технологиясын қолдану арқылы бастапқы деректер қорының сапасын жақсартуға болады.

Кілттік сөздер: онтология, формальді түсініктер анализі, кластерлеу, түсініктер алу, мәліметтер қоры, бутстраппинг

Кіріспе

Онтология белгілі бір білім саласын сипаттайтын спецификацияның формальді ерекшеліктері мен қарым-қатынастарды ресми нақтылау үшін пайдаланылады. Белгілі бір білім саласын ұсыну үшін онтологияның артықшылығы – формальді құрылым [4]. Формальді құрылым – информатика саласында ақпаратты өңдеуді оңайлатады.

Онтология белгілі бір пәндік аймақта білімді жинақтау, қайта пайдалану мүмкіндігі, модель жасау үшін пайдаланылуы мүмкін. Бұл бір пәндік аймақ бойынша мамандарға білімін жақсартуға, жаңа үлгілерді табуға, тәжірибе алмасуға мүмкіндік береді.

Соңғы уақытта онтология құру мәселесіне көптеген зерттеулер жүргізілді. Алайда, онтологияны автоматтандырылған түрде құру мәселесі қазіргі уақытта қанағаттанарлық шешімге ие емес. Белгілі бір пәндік аймақ онтологиясын автоматты түрде құру шешімдерінің бірі – бастапқы кіріс деректер қорына пәндік аймақ мәтіндер корпусын қолданып, Bootstrapping технологиясын концептуалды моделдеу үлгісі ретінде пайдаланып, пәндік аймақ онтологиясының иерархиясын құру.

Пәндік аймақ онтологиясының құрылуы

Қазіргі уақытта онтологиялық инжинеринг саласында онтологияны концептуализациялық ерекшелік ретінде көрсететін екі түрлі полярлық көзқарас бар. Алғашқы көзқарас – онтологияны симуляцияланған субъектілердің қасиеттері мен қарым-қатынастарының дәл математикалық тұжырымдамаларын беруге тырысатын ресми теория ретінде көрсету. Мұндай формальді теориялар, әдетте, математикалық аксиоматтарға негізделген және әртүрлі логикалық формализмдерді пайдаланады, мысалы, предикаттарды есептеу. Екінші көзқарас – онтологияны семантикалық желі ретінде көрсету. Мұндай көзқарастың құрылымын сипаттау өте оңай: оның торабы – пәндік аймақ түсінігі, ал доғалар – таксономикалық қатынастар болып табылатын түсініктер арасындағы байланыстар болып табылады.

Осылайша, екінші типтегі жеңіл онтология – бірінші типтегі ауыр салмақты формальді онтологияға қарағанда елеулі артықшылығы бар: оларды оңай жобалауға болады. Бүгінгі таңда медицина, лингвистикада саласында, электронды коммерцияда және т.б. қазіргі заманғы кең ауқымды пәндік аймақ үшін қолданылатын «өте үлкен ресурстар» жеңіл онтология негізінде құралады [1].

Статистикалық әдістер негізінде пәндік аймақ онтологиясы

Пәндік аймақ онтологиясын моделдеу құрылу жолы, дамыту, ақпаратты өңдеу процесстерінен құрылады. Жалпы пәндік аймақ онтологиясын құру келесідей алгоритмнен тұрады: бастапқы кіріс деректер қорын жинау, үлгі бойынша сөз тіркестерін алу, объект-қатынас кестесін құру, соңғы этапта объектерден түсініктерді алу немесе иерархиясын құру.

Онтологияны құрудың әдістерінің бірі белгілі бір пәндік аймақ үшін объектілерді тікелей жинау болып табылады. Мұндай әдіспен пәндік аймақ онтологиясын құру жолы бастапқыда кіріс деректер қорынан объектілер шығарылып, түсініктер алынады. Түсініктер алынған соң, қайтадан сиректілікті жою үшін толықтырулар жүргізіледі. Сәйкесінше, ұсынылған мақсат бойынша келесі орындалатын жұмыстар мынадай қадамдармен орындалады. Алғашқы бөлімде «формальді анализ түсінінің» негізгі ережелері келтіріледі. Келесі бөлімдерде белгілі бір пәндік аймақ үшін жеңіл салмақты онтология құру алгоритмі сипатталады:

- Формальді түсініктер анализы;
- Пәндік аймаққа қатысты корпус жинау;
- Кіріс деректер қоры негізінде сөйлем ішінен терминдерді алу
- Объект-қасиет кестесінде қуыстарды қалпына келтіру
- Объект-қасиет кестесінен түсініктерді алу

Формальді түсінік анализінің негізгі элементтері

Формальді түсініктер анализі алгебралық тілдік теорияның қолданбалы саласы болып табылады, ал оның шеңберінде алгебра тілінде ұғымдардың түсінігі мен иерархиясын сипаттайтын математикалық формализм ұсынылады [2].

Онтология құру кезінде формальді түсініктер анализін қолданған кезде тек онтология қаңқасын құрастыра аламыз, яғни пәндік аймақ үшін көптеген түсініктердің арасындағы байланыстар анықталады. Мұндай әдіспен объект-қасиет байланысы құрылады. Объект-қасиет байланысы арқылы формальді түсініктер торын құрастыра аламыз. Формальді түсініктер анализінің математикалық негізгі идеясы – кез-келген бинарлы қатынас торын құру және объект-қасиет түсініктерін алу.

Формальді түсініктер анализі үшін пәндік аймақ онтологиясына бастапқы кіріс деректері негізінде объект-қасиет шаблонны қолданылады, мұндағы объект деп пәндік аймақ классын жатқызсақ, қасиет – объектінің қасиеті болып табылады.

Кіріс деректер қоры үшін формальді түсініктер анализі – пәндік аймақ негізінде қасиеттерді бөлуге арналған. Формальді түсініктер анализі деп келесі үштік контекстін айтамыз:

$$K = \langle G, M, I \rangle \quad (1)$$

G – объектілер жиынтығы, M – қасиеттер жиынтығы, I – G мен M нің арасындағы қатынас: gIm дегеніміз, $g \in G$ объектісі $m \in M$ қасиетіне ие болады. Ендігі мәселе кіріс деректер қоры үшін пәндік аймақ негізінде объектілер мен қасиеттердің алыну жолы. Берілген контекст ішінен формальді түсініктердің алынуы формула негізінде көрсетейік.

1-суретте Хассе диаграммасы суреттелген. Ол түсініктердің екі торынан тұрады. Олар өз кезегінде қарастырылып отырған формальді контекст негізінде құрастырылған. Қабылдап алған иерархиялар мен объектілер кластарын анализ жасай отырып осы саланың эксперттері олардың қателіктерін байқайтын еді. Мысалы: креатанин мен мочевины еритін заттар бола тұра, «еритін заттар» түсініктерінен бұрыс шығарылған. Бұл формальді түсініктерді анализдеудің негізгі проблемасы болып табылады егер контекст сиретілген және шектелген (құрамында қуыстар бар) болса, онда алынатын түсініктер және әсіресе олардың иерархиясы бұрыс қалыптасатын болады.

Табиғи тілде сөйлем ішінен құрылымдалмаған сөйлемдерге формальді түсініктер анализі пәндік аймақ онтологиясын құруға қолданамыз. Формальді түсініктер анализін мұндай тәсілмен қолдану модельдеу үшін біртұтас бағыт қолданады. Мұндай бағыт басқа тәсілдерге қарағанда басымдылықты білдіреді.

Пәндік аймаққа қатысты корпус жинақтау

Пәндік аймақ үшін кіріс деректер қоры негізінде «Большой медицинской энциклопедии» сайтынан 13000 медициналық мақаласы алынды [3]. Бастапқы

M. Yemelyanova and O. Baklanova

Mathematical modeling of elastic wave propagation in a homogeneous media 78

S. Zhuzbayev and D. Sabitova

Analysis and synthesis of typical behavioral structures for cyber physical systems' agents 90

D. Zyryanov and N. Denissova

Моделирование систем вентиляции методом теории гидравлических цепей 97

Д. В. Бойков

Статистикалық әдістер негізінде пәндік аймақ онтологиясын құру ... 112

Б.Б. Досанов, М.Е. Мансурова, А.Б. Нугуманова

Современные математические модели, численные методы и пакеты прикладных программ для решения задач фильтрации 121

Б. Жумагулов

Компьютерное моделирование распространения трещины из вершины острого v-образного выреза при антисимметричном изгибе.. 131

В. Кургузов

Использование метода конечных объемов для решения уравнения переноса излучения в прикладных задачах 147

К. Литвинцев

Комплекс некоммерческих программ «SigmaFlow» для решения задач гидродинамики, тепломасообмена и горения 163