

Розглядається застосування технології контент-моніторингу Інтернет-ресурсів для уточнення та поповнення онтології заданої ПрО. Розроблено онтологію «Бездротовий Інтернет в Україні», яка постійно оновлюється й використовується для вирішення задачі вибору оптимального тарифу та Інтернет-провайдерів

Ключові слова: онтологія, контент-моніторинг, бездротовий Інтернет, SWRL-правило, SPARQL-запит

Рассматривается применение технологии контент-мониторинга Интернет-ресурсов для уточнения и пополнения онтологии заданной ПрО. Разработана онтология «Беспроводной Интернет в Украине», которая постоянно обновляется и используется для решения задачи выбора оптимального тарифа и Интернет-провайдеров

Ключевые слова: онтология, контент-мониторинг, беспроводной Интернет, SWRL-правило, SPARQL-запрос

ВИКОРИСТАННЯ ТЕХНОЛОГІЇ КОНТЕНТ- МОНІТОРИНГУ ДЛЯ ВИРІШЕННЯ ЗАДАЧ ОНТОЛОГІЧНОГО ІНЖИНІРИНГУ

Н.В. Рябова

Кандидат технічних наук, доцент,
виконуючий обов'язки завідувача кафедри*

E-mail: ai@kture.kharkov.ua

Ю.В. Сидоренко

тестувальник ПЗ

Компанія Injoit Ltd, Харківський офіс
вул. Серпова, 4, м. Харків, Україна, 61166

М.В. Білоіваненко

Науковий співробітник*

*Кафедра штучного інтелекту

Харківський національний університет радіоелектроніки
пр. Леніна, 16, м. Харків, Україна, 61166

Вступ

Постійне зростання обсягів інформації, доступної завдяки Інтернет-джерелам, стало передумовою таких проблем як: непропорційне зростання інформаційного шуму через слабку структурованість інформації; багаторазове дублювання інформації; невідповідність формально релевантної інформації дійсним потребам її споживачів; поява паразитної інформації та спаму. Наслідком цих проблем стало інформаційне перевантаження з одного боку, та нестача необхідних знань з іншого, які відчувають Інтернет-користувачі. Традиційні інформаційно-пошукові системи часто виявляються нездатними задовольнити запити користувача. Охоплення, узагальнення великих динамічних інформаційних масивів, які безперервно генеруються в мережі, вимагає якісно нових підходів. Виникла необхідність створення методів моніторингу інформаційних ресурсів, тісно пов'язаних з методологією контент-аналізу.

Контент-моніторинг – це змістовний аналіз інформаційних потоків з метою отримання необхідних якісних і кількісних зрізів, який проводиться безперервно у часі. Методи контент-моніторингу – це адаптація класичних методів контент-аналізу до умов динамічних інформаційних масивів, наприклад, потоків інформації з Інтернет. Типове завдання контент-моніторингу – побудова діаграм динаміки появи понять у часі. Системи моніторингу та контент-аналізу забезпечують регулярний пошук і «скачування» інформації по заданим темам і з заданих сайтів, а також аналіз змі-

сту «скачаних» документів. Це масштабовані системи, в яких реалізовані різні математичні та лінгвістичні алгоритми аналізу текстових даних. Вони мають розвинуті графічні інтерфейси, багаті можливості візуалізації та маніпулювання з даними, надають доступ до різних джерел даних, функціонують в архітектурі клієнт-сервер [1].

Для семантичного структурованого представлення інформації, яка є результатом видобування з Інтернет-джерел засобами систем контент-моніторингу найбільш перспективним є онтологічний підхід. На формальному рівні, онтологія – це система, що складається з набору понять і набору тверджень про ці поняття, на основі яких можна будувати класи, об'єкти, відносини, функції і теорії; предметна область (ПрО) розглядається як сукупність об'єктів і відносин між ними, реалізована у мережевій моделі знань.

Дані технології активно застосовуються в багатьох ПрО, але досить багато залишається не вивченими і не автоматизованими. Прикладом можуть послужити актуальні на сьогоднішній день технології бездротового Інтернет. Бездротові технології дозволяють незалежно від місця знаходження бути підключеними до мережі, обмінюватися даними, отримувати необхідну інформацію з Інтернет або просто спілкуватися. Велика кількість Інтернет-провайдерів і послуг, що надаються ними, породила проблему вибору найбільш вигідних тарифів користувачами відповідно до заданих параметрів.

Для вирішення даної проблеми за допомогою системи контент-моніторингу будуть відстежуватися

всі зміни і оновлення, які публікуються на Інтернет-сторінках, і потім структуруватися. Таким чином, розроблена експертом онтологія Про буде постійно оновлюватися та доповнюватися необхідною актуальною інформацією. Даний підхід дуже зручний для подальшого порівняння всіх тарифів, що цікавлять користувача, і вибору найбільш вигідного, оскільки користувач матиме змогу вказати критерій переваги, за яким буде відбуватися порівняння тарифів.

1. Аналіз предметної області та постановка задачі дослідження

Сьогодні доступ до динамічної складової Інтернет – новинним ресурсам – ускладнений. Різноманітність інформації, в тому числі і новинних повідомлень, у мережі не може бути корисною на практиці за відсутності ефективного доступу [1]. Тобто, вся мережа Інтернет орієнтована на показ користувачам окремих сайтів і погано пристосована для автоматизованого збору інформації, її класифікації та аналітичної обробки. Сьогодні представлення інформації на різних сайтах настільки відрізняється за оформленням та розташуванням, що відбирати її і обробляти можна тільки вручну. Оптимальне рішення, здатне допомогти орієнтуватися в новинній інформації Інтернет, сьогодні надають інформаційні служби нового типу – системи синдикації новин. Під синдикацію розуміється збір інформації в Інтернет і подальше поширення її фрагментів відповідно до потреб користувачів. Крім того, служби синдикації забезпечують публікацію одних і тих же даних на різних сайтах, у тому числі, призначених для кишенькових комп'ютерів і мобільних телефонів. Для вирішення завдання синдикації новин було створено кілька форматів опису даних на основі XML. Найпоширеніший формат отримав назву RSS, що означає Really Simple Syndication, Rich Site Summary, хоча спочатку він називався RDF Site Summary. Зміст всіх цих аббревіатур полягає в простому способі узагальнення та розподілу інформаційного наповнення web-сайтів – синдикації контенту.

Служби синдикації забезпечують одночасну публікацію одних і тих же даних на різних сторінках, сайтах і мобільних пристроях (у тому числі, в кишенькових комп'ютерах і мобільних телефонах), а також доставку інформації користувачам [1,2]. Технологія синдикації Інтернет-новин включає “навчання” програм збору інформації структурі вибраних джерел, сканування інформації, її нормування, приведення до загального формату (останнім часом все частіше до XML, RSS), класифікацію, кластеризацію і доставку користувачам різними каналами (e-mail, WWW, Wap, SMS та ін.).

Формат RSS, що отримав справжнє широке поширення, дійсно забезпечує погоджений спосіб резюмувати зміст web-сайтів. Аббревіатура RSS припускає неоднозначні, але понятійно близькі трактування – Really Simple Syndication, Rich Site Summary, RDF Site Summary (RDF – Resource Description Framework, стандарт метаданих, тісно пов'язаний з XML). Спочатку RSS створювався компанією Netscape для свого порталу Netcenter як один з перших XML-додатків,

але потім швидко завоював популярність і став досить широко використовуватися.

Після того, як інформація перетворена у формат RSS, створену інформаційну стрічку можна читати за допомогою спеціальних програм або імпортувати дані з неї на web-ресурси. Сьогодні практично усі ведучі світові інформаційні сайти, “живі журнали”, що працюють в Інтернеті, використовують RSS в якості інструменту оперативного представлення свіжої інформації.

Різновидом використання RSS є новинні фіди (feed). Фід – це файл у форматі RSS, в який записується новинна інформація web-ресурса(точніше, анонси оновлень). Якщо є необхідність оперативно відстежувати зміни на сайті, використовуючи фід, що їх містить, то можна робити це за допомогою программы-агрегатора не відвідуючи самого сайту. Системи синдикації Інтернет-новин вирішують проблему пошуку необхідної інформації, але залишають без уваги такі завдання як узагальнення даних, їх обробку і аналіз.

Одним з найбільш перспективних напрямків узагальнення інформаційних потоків в даний час є контент-моніторинг. Його поява була викликана, перш за все, завданнями систематичного відстеження тенденцій і процесів у інформаційному середовищі, що постійно оновлюється інформаційному середовищі. Під контент-моніторингом найчастіше розуміють змістовний аналіз інформаційних потоків з метою отримання необхідних якісних і кількісних зрізів, який ведеться постійно протягом не визначеного заздалегідь проміжку часу. Методи контент-моніторингу – це адаптація класичних методів контент-аналізу до умов динамічних інформаційних масивів, наприклад, потоків інформації з Інтернет.

Сутність контент-моніторингу полягає в поєднанні змістовних та кількісних методів аналізу текстів. Послідовність аналізу проблеми, що досліджується, умовно можна поділити на такі стадії: змістовний (якісний) аналіз сукупності текстів; формалізований (кількісний) аналіз інформаційних масивів. Механізм обробки контенту виконує наступні функції: перетворення зібраної інформації до обраного внутрішньо системного формату; систематизацію інформації; вибіркового розподілу інформаційних повідомлень за запитами користувачів; завантаження та індексацію оперативних і ретроспективних баз даних. Контент-аналіз – це систематична числова обробка, оцінка та інтерпретація форми та змісту інформаційного джерела.

Це якісно-кількісний метод вивчення документів, який характеризується об'єктивністю висновків і строгістю процедури, і являє собою квантифікаційну обробку тексту з подальшою інтерпретацією результатів.

У автоматизованій технології контент-моніторингу існує кілька важливих особливостей: використання ключового фрагмента публікації як одиниці формування текстового інформаційного масиву; формування банку ключових фрагментів публікацій є об'єднанням двох взаємозалежних автоматизованих процесів: аналітико-синтетичної переробки та багаторівневої процедури контент-аналізу текстів публікацій; індексація ключових фрагментів пу-

блікацій відбувається за допомогою багатомасової класифікації.

Унікальність методології контент-моніторингу полягає у тому, що вона не прив'язана ні до конкретної СУБД, ні до конкретних видів джерела інформації, ні до тематики інформації. Таким чином, в простому вигляді ідею контент-моніторингу можна сформулювати як постійне виконання вузько окресленого своїми завданнями контент-аналізу безперервних інформаційних потоків. Саме безперервне відтворення в часі процесу обробки вхідних даних є найхарактернішою особливістю контент-моніторингу. Контент-моніторинг дозволяє автоматично формувати доступні для огляду дайджести, замінюючи ними неозорі обсяги електронних ЗМІ. Базою для побудови технології контент-моніторингу є механізми виділення найбільш важливих термів – слів і словосполучень, пов'язаних між собою змістовно і відповідаючих тематиці, яку задає користувач. Вага термів визначається на основі статистичних методів аналізу текстів з урахуванням емпіричних критеріїв, які уточнюються в процесі роботи. Один із практичних результатів контент-моніторингу, разом з відомими соціологічними висновками, автоматичне формування дайджестів, які у лаконічній формі відображають зміст відібраних тематичних інформаційних каналів.

Впровадження технології контент-моніторингу забезпечує такі переваги: одержання оперативних кількісних і якісних аналітичних зрізів; включення робочих місць співробітників-аналітиків у динамічний інформаційний простір, своєчасне надання необхідної профільної інформації; забезпечення цілеспрямованої роботи співробітників, усунення факторів відволікання уваги, властивих Інтернет; захист даних, конфіденційність. Завдяки таким характеристикам використовуваних інструментальних засобів роботи з інформаційними потоками, як оперативність, повнота і релевантність, а також наявність єдиного захищеного інтерфейсу, технологія контент-моніторингу може сприяти значному підвищенню ефективності та якості інформаційно-аналітичної роботи.

Задля реалізації контент-моніторингу у даній роботі передбачається використання системи InfoStream [2], розробленої в інформаційному центрі “ЕЛВІСТІ” та призначеної для автоматизованого збору інформації з відкритих web-сайтів, її обробки, систематизації, узагальнення і забезпечення доступу до неї. Ядром системи обробки контенту є повнотекстова інформаційно-пошукова система InfoReS, що забезпечує обробку даних в трьох основних режимах: виборчого поширення; інтерактивного доступу до повнотекстових баз даних; контент-моніторингу. Користувачі можуть отримати доступ до даних у форматі RSS за допомогою спеціальних програм, званих RSS-агрегаторами. Програма-агрегатор дозволяє збирати усі публікації з тих, що цікавлять користувача сайтів разом, одночасно стежити за появою новин на усіх сайтах відразу і читати їх короткий зміст, не відвідуючи кожен цей сайт окремо.

Методи контент-моніторингу – це адаптація класичних методів контент-аналізу до умов динамічних інформаційних масивів, наприклад потоків інфор-

мації з Інтернет. На відміну від систем інтеграції інформації, які реалізують ідею збору і накопичення усієї доступної інформації, як з внутрішніх, так і з зовнішніх джерел, системи контент-моніторингу дозволяють виявляти неочевидні закономірності в документальних масивах даних або текстів – так звані латентні (приховані) знання. Таким чином, з масиву отриманих повідомлень/документів система відокремлює найбільш релевантні запити для подільшої їх обробки експертом, який за допомогою середовища Protégé (або TopBraid) оновлює вже існуючу онтологію або додає в неї нові дані, таким чином, роблячи інформацію, що знаходиться в онтології, актуальною. Постійно зростаюча кількість Інтернет-провайдерів і інформація про тарифи, часто оновлюється в Інтернет-джерелах, робить непростим самостійний вибір пропозицій користувачами.

Основним завданням, що вирішується в роботі, є побудова та оновлення онтології Про «Бездротовий Інтернет в Україні» на підставі інформації, отриманої засобами контент-моніторингу. За допомогою онтологічного підходу можливо наглядно і доступно представити всю інформацію про існуючі технології бездротового Інтернет, Інтернет-провайдерів і діючих тарифах як деяку структуру. Всі існуючі технології, а саме 2G, 3G і 4G, будуть концептами, які будуть включати в себе всі технології, які їм відповідають і Інтернет-провайдерів, які їх використовують. Для кожного провайдера бездротового Інтернет будуть вказані їх характеристики, такі як швидкість передачі даних, швидкість завантаження, діапазон покриття, тарифи та безлімітні пакети. Дана структура буде представляти ринок бездротового Інтернет в повному обсязі, що дозволить легко орієнтуватися в представниках і їх послугах.

Надалі, для того, щоб вибрати найбільш вигідний тариф, в роботі будуть створені SWRL-правила, за допомогою яких система буде порівнювати окремі тарифи за критерієм переваги, який оберє для себе користувач. Це будуть властивості екземплярів класів. Далі, коли користувач буде відвідувати сайт для вибору тарифу, він повинен вказати критерій переваги, який буде активувати те чи інше SWRL-правило, що відповідає кожній властивості окремо. Відповідно до всіх тарифів, що відповідають правилу, необхідно розрахувати такий, що буде якнайкраще відповідати потребам користувача. Для визначення величини відповідності тарифів потребам необхідно розрахувати вагу кожного тарифу за формулою, яку обирає експерт самостійно. В результаті оцінка кожного тарифу буде варіюватися від 0 до 1 відповідно.

2. Порівняльний аналіз стану розробки бездротових технологій та відповідних послуг Інтернет-провайдерів

Під поняттям бездротовий Інтернет розуміють велику кількість різних технологій, які дозволяють отримувати доступ в Інтернет без дротових з'єднань. Бездротові технології дозволяють нам, де б ми не знаходилися, бути підключеними до мережі, обмінюватися даними, знаходити потрібну інформацію в Інтернеті або ж просто спілкуватися з кимось. Сьогодні вже очевидно, що створення суцільної зони

покриття бездротового сигналу в різних стандартах – вимога часу, і динаміка поширення бездротових технологій стає все більш стрімкою. Купуючи послуги бездротового Інтернету можна орієнтуватися на наступні важливі умови: доступність; швидкість роботи; простота використання; ціна. Сьогодні існує ряд можливостей бездротового підключення переносного комп'ютера до Інтернет, кожна з яких має свої плюси і мінуси. Умовно його можна поділити на шість поколінь 0G, 1G, 2G, 3G, 4G і 5G. В окрему категорію можна віднести супутниковий Інтернет і Wi-Fi.

2.1. Покоління бездротових технологій зв'язку

Промисловість бездротового зв'язку почала створення, розвиток і еволюцію їх технологій на початку 1970-х років. В останні кілька декад, бездротові мобільні технології розвинулись до четвертого і п'ятого поколінь технологій, названих від 0G до 5G. Поняття мобільності було запропоновано в 1G-технології, що зробило можливим великомасштабний бездротовий зв'язок. Цифровий зв'язок змінили аналогові технології у 2G, які значно поліпшили якість бездротового зв'язку. Передача даних, як додаток до голосового зв'язку, була у центрі уваги у 3G-технологіях та конвергентних мережах як для передачі голосу, так і даних.

Бездротовий зв'язок починався з того, що можна було б назвати 0G. Ці ранні системи мобільного телефонного зв'язку можуть бути виділені з раніше закритих радіотелефонних систем в тому, що вони були доступні як комерційні служби, які були частиною громадської комутованої телефонної мережі, з їх власними телефонними номерами, а не частиною закритої мережі, такі як поліція, радіо або диспетчерська система таксі. 1G технологія замінила 0G технологію, яка показала мобільні радіо телефони і такі технології, як мобільні телефонні системи (MTC), розширені мобільні телефонні системи (AMTC), поліпшена мобільна телефонна служба (IMTC), і натиск для спілкування (PTT). Його послідовник, 2G, побудований на використанні цифрових сигналів, а бездротові мережі 1G використовують аналогові радіо сигнали. 2G-технології можуть бути розділені на Time Division Multiple Access (TDMA) та Code Division Multiple Access (CDMA) стандарти в залежності від типу багаторазового використання. 2G використовує CODEC (алгоритм компресії-декомпресії) для стиснення і мультиплексування цифрових голосових даних. Завдяки цій технології 2G-мережа може вмістити більше дзвінків за пропускну здатність, як і 1G-мережа.

3G-технології надають мобільним операторам можливість пропонувати користувачам широкий спектр більш передових сервісів і в той же час досягати більшої ємності мережі шляхом поліпшення спектральної ефективності. Сервіси включають широкомасштабну бездротову голосову телефонію, відео дзвінки, широкосмугові безпроводні передачі даних, все в мобільному середовищі. На відміну від IEEE 80-2.11 (так званого Wi-Fi або WLAN) мереж, 3G-мережі – це широкомасштабні стільникові телефонні мережі, які включають в себе високошвидкісний доступ в Інтернет і телефонію. 4G належить до четвертого покоління стільникових бездротових стандартів. Вона

є наступницею 3G і 2G наборів стандартів. Головним чином, це розширення 3G-технології з більшою пропускну здатністю і послугами, запропонованими у 3G. У 4G очікується висока якість аудіо/відео потоку від початку до кінця Інтернет-протоколу. 5G (п'яте покоління мобільних мереж або п'яте покоління бездротових систем) – назва, що використовується у деяких дослідницьких роботах і проектах, щоб позначити наступний важливий етап мобільних стандартів зв'язку, які, як очікується, будуть закінчені у період між 2011 та 2013 роками.

Wi-Fi – найпоширеніша в Україні технологія швидкісної бездротової передачі даних, на неї припадає близько 80% ринку радіо доступу. По-перше, Wi-Fi підтримується всіма виробниками мережевого бездротового обладнання, що позитивно впливає на розвиток цього стандарту. По-друге, в більшості країн таке обладнання не потребує отримання ліцензії. По-третє, постійно зростає швидкість передачі даних в Wi-Fi. Незабаром з'являться досконаліші стандарти Wi-Fi, де швидкість передачі даних досягне 1Гбіт/с, але і це не межа. Крім того, підтримка стандарту Wi-Fi з недавнього часу встановлюється на мобільних телефонах. Практично всі операційні системи (Windows, Linux, Mac OS та ін) підтримують даний стандарт. Також цей стандарт підтримують операційні системи для мобільних пристроїв (Windows Mobile, Symbian, Android, Palm OS), що дає можливість розробникам впроваджувати безліч сервісів, додатків, ігор і т.д.

Однак, у стандарті Wi-Fi є і ряд недоліків. Найголовніший недолік – обмежений радіус покриття. Стандартний маршрутизатор для будинку із застосуванням всенаправленої антени має радіус дії, в середньому, 30-50 метрів в приміщенні і близько 100 метрів зовні. Також на радіус дії стандарту Wi-Fi істотно впливають різні електроприлади, звичайні дзеркала, опади у вигляді дощу. Ці фактори впливають і на швидкість передачі даних. Щоб покрити значну територію в умовах міста необхідно створити величезну кількість точок доступу, радіус покриття яких при несприятливих умовах може зменшуватися. Варто зауважити, що безпека даних, переданих через протокол Wi-Fi залишає бажати кращого. Тому багато організацій використовують додаткове шифрування для захисту від несанкціонованого вторгнення.

Супутниковий Інтернет – спосіб забезпечення доступу до мережі Інтернет з використанням технологій супутникового зв'язку (як правило, в стандарті DVB-SaBo DVB-S2). Існує два способи обміну даними через супутник: односторонній (one-way), іноді називають «асиметричним», коли для прийому даних використовується супутниковий канал, а для передачі – доступні наземні канали; двосторонній (two-way), іноді називають «симетричним» - коли для прийому для передачі використовуються супутникові сигнали. Двосторонній супутниковий Інтернет є дуже якісним, оскільки дозволяє досягати великих швидкостей при передачі та відправці, але він є досить дорогим і потребує отримання дозволу на радіопередавальне устаткування. Висока вартість двостороннього Інтернету виявляється повністю виправданою за рахунок в першу чергу набагато більш

надійного зв'язку. На відміну від одностороннього доступу, двосторонній супутниковий Інтернет не потребує ніяких додаткових ресурсів.

2.2. Провайдери бездротового Інтернет

Технології бездротового Інтернет стрімко розвиваються. Десятки пропозицій від відомих, і не дуже компаній, як правило, ставлять у глухий кут людину, робота якої не пов'язана з високими технологіями. Основними провайдерами бездротового Інтернет з використанням відповідних технологій є такі компанії:

CDMA UKRAINE; Utel, Укртелеком; People net; Kyivstar; MTC; Інтертелеком; Beeline (3G);

Alternet; FreshTel; IMC; Intelcom; Pan Telecom; UNTC; АП-ТРЕЙДІНГ; Ведекон(4G);

супутниковий інтернет: SkyDSL (Скай ДСЛ).

Компанія CDMA UKRAINE – український оператор мобільного зв'язку, що надає послуги мобільного зв'язку і бездротового доступу в Інтернет на стандарті CDMA. Працює в діапазоні частот 800 МГц. Швидкість завантаження інформації до 3,1 Мбіт/сек. Компанія CDMA UKRAINE належить американо-українській компанії ITC – 100% акцій. Мережа зв'язку Utel працює за технологією UMTS/WCDMA. Utel постійно розвивається, збільшуючи кількість міст та абонентів. Переваги Utel: покриття UMTS-2100 є в усіх великих містах та обласних центрах; можливість протестувати якість надаваних послуг не підписуючи контракти; наземні канали зв'язку мають велику пропускну здатність (Utel – філія Укртелекому, який у свою чергу має велику кількість оптоволоконних ліній по всій території України); серед всіх 3G операторів показує найвищу швидкість на прийом.

PEOPLEnet – один з лідерів швидкісного мобільного Інтернет в Україні. Використовує технологію 3G 1x EV-DO Revision A. Швидкість роботи до 3,1 Мбіт/сек. Для роботи в Інтернеті PEOPLEnet пропонує: USB модеми; PCMCIA-карти; мобільний телефон з модемом. Київстар – мобільний оператор, один з лідерів ринку мобільного зв'язку. Надає своїм абонентам широкий спектр послуг:

– EDGE (Enhanced Data for Global Evolution) – дає можливість відправляти й одержувати великі обсяги інформації за допомогою мобільних телефонів за лічені секунди. Швидкість EDGE до 384 Кбіт/с, а це в 8-10 разів більше можливостей GPRS. Мобільні телефони налаштовувати не треба, телефон автоматично перейде в режим EDGE, як тільки потрапить в зону покриття EDGE;

– GPRS (General Packet Radio Service) – технологія передачі даних в GSM-мережах.

Можливості GPRS дозволяють: зменшити час з'єднання з мережею Інтернет; збільшити швидкість прийому / передачі інформації в 3-4 рази; приймати виклики, а також телефонувати, не розриваючи при цьому доступу до мережі Інтернет; використовувати нову, вигідну систему тарифікації доступу до Інтернету. Для доступу до Інтернет за допомогою послуги WAP або традиційного доступу до мережі Інтернет, потрібно налаштувати доступ.

MTC – оператор мобільного зв'язку, який охоплює близько 97% території всієї України, надає такі послуги як:

– "Wi-Fi від МТС". Користуватися можна у будь-якому місці, де є покриття хот-спот цього провайдера (готель, бізнес-центр тощо). Швидкість до 50 Мбіт / сек;

– МТС Коннект 3G. Середня швидкість – від 200 до 800 Кбіт /с (максимальна до 3.1 Мбіт / с). Для надання послуги 3G необхідно підписати контракт і провести активацію. Технологія CDMA 2000 1x EVDO REV.A – 3G. Дана послуга доступна більш ніж 150 містах України.

Інтертелеком – перший національний CDMA оператор. Пропонує послуги: мобільного зв'язку за стандартом CDMA; безпроводний доступ в Інтернет (21 регіоні України). Компанія Alternet надає послуги WiMAX. Радіотехнологія WiMAX здатна працювати без перешкод і втрати якості сигналу в будь-яку погоду. Мережа Alternet використовує стандарт 802.16d, (фіксований стандарт WiMAX). Тому стабільний сигнал забезпечується тільки на фіксованих об'єктах.

Послуги компанії Alternet: Інтернет-доступ до 4 Мбіт /с; корпоративні мережі; банківські мережі; VPN рішення; відеоспостереження (3 дні); IP телефонія; Wi-Fi рішення.

FreshTel – це перший провайдер високошвидкісного бездротового Інтернету 4G в Україні, а також перший міжнародний оператор MobileWiMAX в Україні, у мережі якого почали роботу ноутбуки та нетбуки із вбудованим WiMAX адаптером від Intel. Компанія Intelcom має представництва у 23 містах (80 населених пунктах).

Послуги: доступу до мережі Інтернет; передачі даних; віддалений моніторинг; відео спостереження; підключення систем банкоматів; устаткування Motorola, Cisco і Dell; швидкість з'єднання – до 40 Мбіт / с. SkyDSL – це провайдер супутникового Інтернету. Працює за принципом однобічної схеми передачі даних (приймальний канал). Спосіб підключення до інтернету: Dial-UP, GPRS, ADSL, Cable modem. Для організації послуг потрібно супутникова (параболічна) антена, яка налаштована на супутник Eutelsat-1, Turksat 2A, Turksat 2B. Швидкість передачі із супутника на «тарілку» досягає 16 Мб/с, це залежить від завантаженості. Присутня велика асиметрія з'єднання, на передачу швидкість до 56 Кбіт/с, а на прийом – до 16 Кбіт/с.

3. Побудова онтологічної моделі предметної області «Бездротовий Інтернет в Україні»

Найбільш загальний спосіб представлення знань, при якому предметна область розглядається як сукупність об'єктів і відносин між ними, реалізований у мережевій моделі знань. Одним з головних досягнень консорціуму по всесвітній павутині (W3C) є розробка стандарту опису онтологій. Найбільш визнаним є визначення онтології, згідно з яким онтологія є точною специфікація концептуалізації. Будь-яка онтологія передбачає виділення її трьох взаємопов'язаних компонентів: таксономії термінів, описів сенсу термінів, а також правил їх використання та обробки. Онтологія використовується для структурування інформації та знань. До основних

завдань, що вирішуються за допомогою онтології, відносяться: створення і використання баз знань, організація семантичного пошуку; реалізація механізмів міркувань, уявлення сенсу в метаданих про інформаційний ресурс.

Інформаційні системи різного призначення на основі онтологій на практиці показали свою ефективність, проте побудова онтології вимагає експертних знань у досліджуваній ПрО. Серед інструментальних засобів для підтримки розробки онтологій найбільш поширеними є Protégé та Top Braid Composer, які використовуються в даній роботі. Але на даний момент не існує повноцінних інструментів автоматичної побудови предметних онтологій, хоча ведуться роботи в цьому напрямі. Для первісного добування інформації з різних джерел у даний час використовуються підходи, засновані на автоматизації видобування знань з даних, текстів і веб-ресурсів (data-mining, text-mining, web-mining), однак отримані дані зазвичай вимагають їх обробки експертом ПрО [3,4].

В даній роботі продовжено дослідження, розпочаті в [5], щодо розробки онтології, яка описує галузь бездротового Інтернет в Україні та включає в себе: технології бездротового Інтернет; провайдерів Інтернет; послуги та тарифи провайдерів, їх характеристику. Для створення онтології був вибраний низхідний процес, тобто розробка починається з визначення загальних понять предметної галузі із подальшою її конкретизацією.

Опишемо стисло основні складові побудованої онтологічної моделі.

3.1. Класи та підкласи онтології

У центрі більшості онтологій знаходяться класи, які описують поняття ПрО. В даній роботі для опису бездротового Інтернет було створено п'ять класів: Tariff (тарифи), User_Prefer (переваги користувачів), Utility (вигідність), Wireless_Internet_Service_Provider (провайдери бездротового Інтернет), Wireless_technology (бездротові технології):

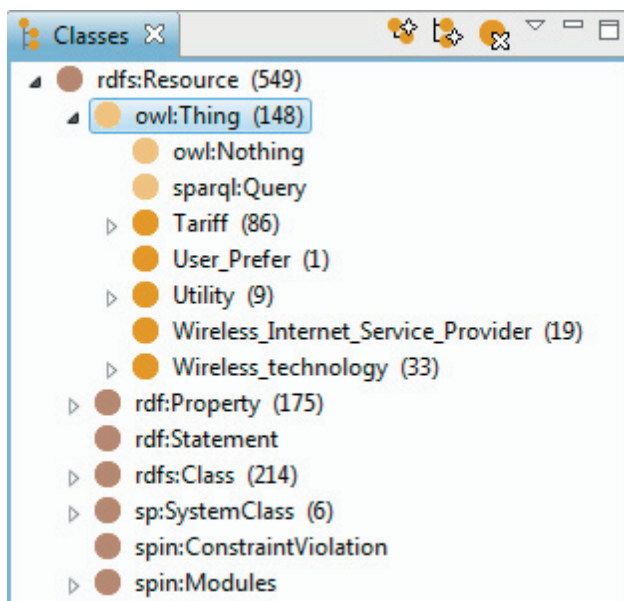


Рис. 1. Класи онтології «Бездротовий Інтернет в Україні»

Клас може мати підкласи, які представляють більш конкретні поняття, ніж надклас. Для класу Tariff було створено три підкласи: Mobile_tariff, Satellite_tariff, Wi-fi_tariff:



Рис. 2. Підкласи класу Tariff

Для класу Utility було створено два підкласи Satellite та Wireless:



Рис. 3. Підкласи класу Utility

Для класу Wireless_technology підкласами було створено три класи: Mobile_technology (мобільні технології), Satellite_Internet (супутниковий Інтернет) та Wi-Fi:

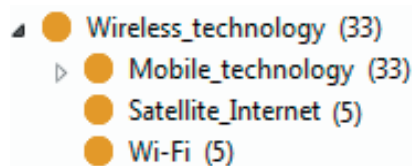


Рис. 4. Підкласи класу Wireless_technology

Клас Mobile_technology також включає в себе підкласи, що є поколіннями мобільного Інтернет: 0G, 1G, 2G, 2.5G, 2.75G, 3G, 3.5G, 3.75G, 4G та 5G:

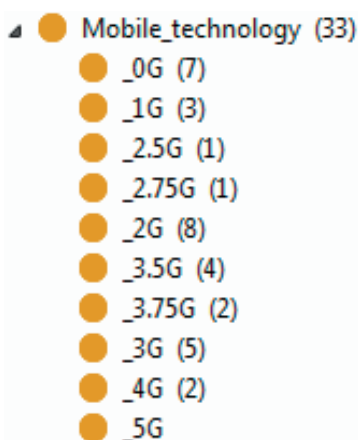


Рис. 5. Підкласи класу Mobile_technology

Кожен із підкласів Mobile_technology містить у собі всі технології, які відносяться до певного покоління та які є екземплярами відповідних класів.

Наприклад, екземплярами класу 2G є наступні технології:

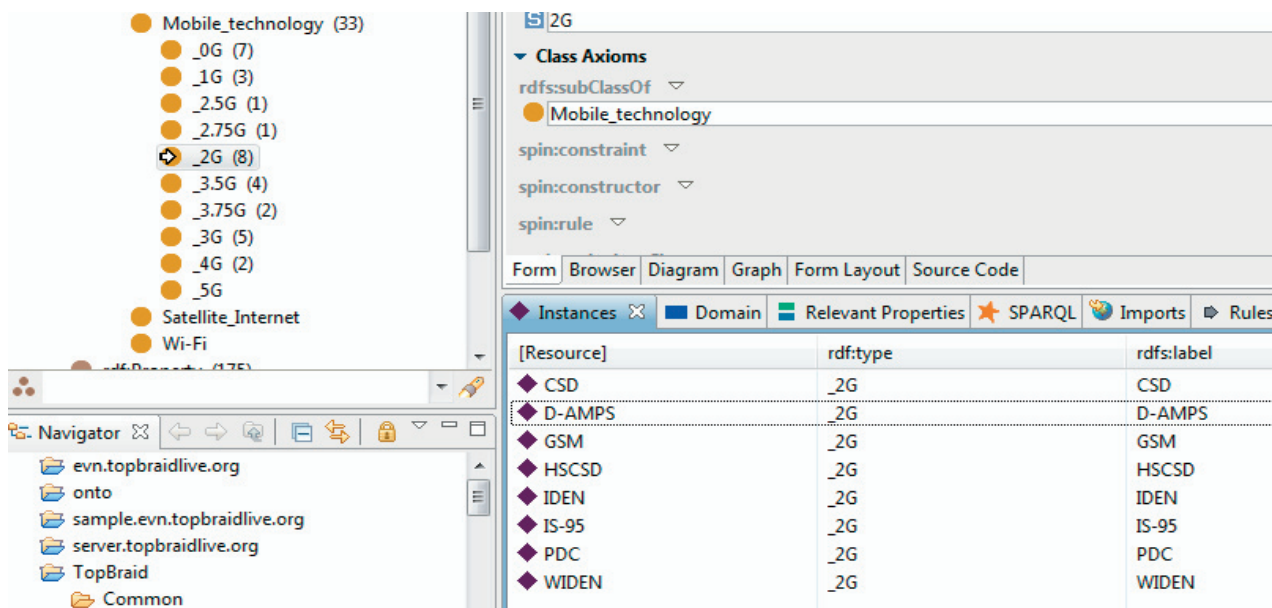


Рис. 6. Екземпляри класу 2G

3.2. Властивості класів

Класи самі по собі не надають достатньо інформації для відповіді на питання перевірки компетентності. Після визначення певної кількості класів необхідно описати внутрішню структуру понять. Було вибрано класи зі складеного списку термінів для опису предметної галузі, які уже було створено. Більшість решти термінів будуть властивостями цих класів. Для кожної властивості зі списку необхідно визначити, який клас вона описує. Ці властивості стануть слотами, що прив'язані до класів. Усі підкласи класу наслідують слот цього класу. Слот повинен бути прив'язаний до найзагальнішого класу, у якого може бути дана властивість. Слоти описують властивості класів та екземплярів. У роботі були створені наступні властивості екземплярів:

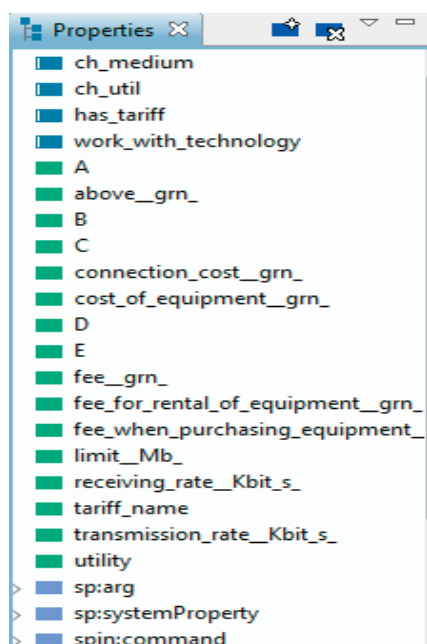


Рис. 7. Властивості екземплярів класів

До властивостей типу owl:ObjectProperty відносяться наступні:

а) work_with_technology. Властивість, що притаманна кожному провайдеру бездротового Інтернет і вказує з якою технологією він працює:



Рис. 8. Зв'язок, встановлений для властивості work_with_technology

б) has_tariff. Властивість, яка встановлює зв'язок між провайдером та тарифом і вказує, які тарифи надає кожен провайдер користувачам:

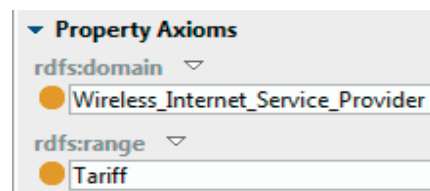


Рис. 9. Зв'язок, встановлений для властивості has_tariff

в) ch_util. Властивість, що встановлює зв'язок між класами User_Prefer та Utility, таким чином, що для кожного користувача можна визначити, чому саме він надає перевагу:

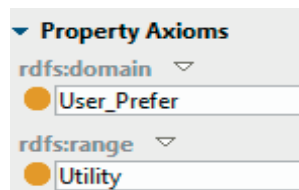


Рис. 10. Зв'язок, встановлений для властивості ch_util

d) `ch_medium`. Властивість визначає, який саме тариф обирає користувач:

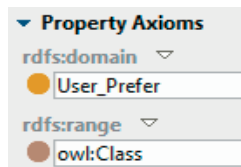


Рис. 11. Зв'язок, встановлений для властивості `ch_medium`

Також визначені властивості типу `owl:DatatypeProperty`:

а) А, В, С, D, Е – визначаються для класу `Utility` та являються скороченням для запису певної властивості у формулу розрахунку критерію переваги. Наприклад:

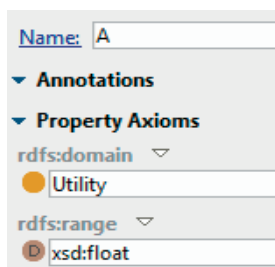


Рис. 12. Характеристики властивості А

b) `above_(grn)`. Властивість, яка вказує вартість кожного використаного мегабайту поверх трафіка.

c) `connection_cost_(grn)`. Вартість підключення послуг бездротового Інтернет.

d) `cost_of_equipment_(grn)`. Властивість супутникового Інтернет, яка відповідає вартості устаткування, указану в гривнях.

e) `fee_(grn)`. Абонентська плата.

f) `fee_for_rental_of_equipment_(grn)`. Вартість, яку сплачує користувач за послуги супутникового Інтернет при аренді устаткування.

g) `fee_when_purchasing_equipment_(grn)`. Вартість, яку сплачує абонент після придбання устаткування.

h) `limit_(Mb)`. Трафік, указаний в мегабайтах.

i) `receiving_rate_(Kbit/s)`. Швидкість завантаження.

j) `transmission_rate_(Kbit/s)`. Швидкість передачі.

k) `tariff_name`. Назва тарифного плану.

l) `utility`. Критерій вигідності.

Останнім кроком у розробці онтології є створення окремих екземплярів класів в ієрархії. Для визначення окремого екземпляра класу необхідно:

– вибрати клас;

– створити окремий екземпляр цього класу;

– ввести значення слотів.

До класу `Mobile_tariff`, що вказує на тип тарифів, який надають провайдері бездротового Інтернет належать наступні екземпляри:

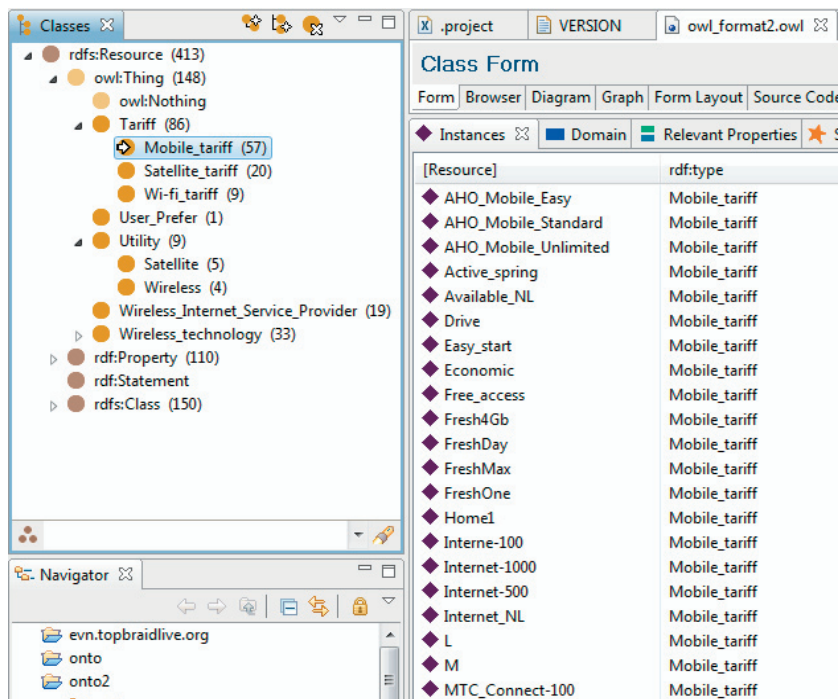


Рис. 13. Екземпляри класу `Mobile_tariff`

До класу `Satellite_tariff`, що вказує на те, що тарифи провайдерів належать до класу супутникових, належать наступні екземпляри:

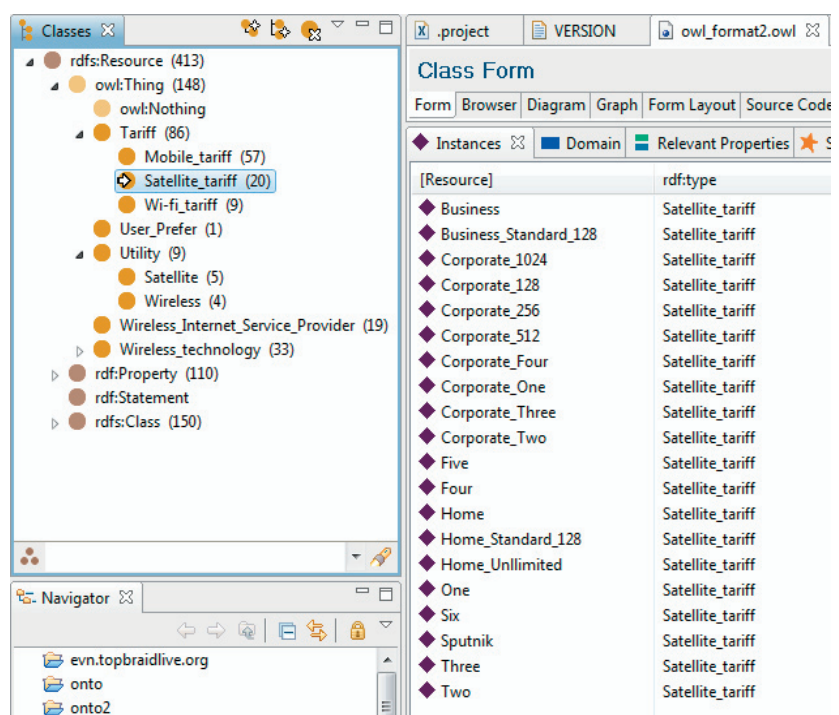


Рис. 14. Екземпляри класу `Satellite_tariff`

До класу Wi-fi_tariff, що вказує на тип тарифу Wi-fi, належать наступні екземпляри:

The screenshot shows the Protege interface. On the left, the 'Classes' browser displays a hierarchy: **rdfs:Resource** (413) > **owl:Thing** (148) > **Tariff** (86) > **Wi-fi_tariff** (9). The **Wi-fi_tariff** class is selected. On the right, the 'Class Form' tab is active, showing a table of instances.

[Resource]	rdf:type
◆ AHO_Wi-fi_Ulimited	Wi-fi_tariff
◆ AHO_Wifi_10Gb	Wi-fi_tariff
◆ _0401	Wi-fi_tariff
◆ _0403	Wi-fi_tariff
◆ _0405	Wi-fi_tariff
◆ _0410	Wi-fi_tariff
◆ _1	Wi-fi_tariff
◆ _107235	Wi-fi_tariff
◆ _7755	Wi-fi_tariff

Рис. 15. Екземпляри класу Wi-fi_tariff

До підкласу Wireless класу Utility входять наступні екземпляри:

The screenshot shows the Protege interface. On the left, the 'Classes' browser displays a hierarchy: **rdfs:Resource** (413) > **owl:Thing** (148) > **Utility** (9) > **Wireless** (4). The **Wireless** class is selected. On the right, the 'Class Form' tab is active, showing a table of instances.

[Resource]	rdf:type	rdfs:label
◆ Fee	Wireless	$= A*0.5 + B*0.003 + C*0.2 + D*0.1$
◆ Limit	Wireless	$= B*0.5/100 + A*0.3 + C*0.2 + D*0.1$
◆ Receiving_rate	Wireless	$= C*0.5 + D*0.3 + A*0.2 + B*0.1/100$
◆ Transmission_rate	Wireless	$= D*0.5 + C*0.3 + A*0.2 + B*0.1/100$

Рис. 16. Екземпляри класу Wireless

До підкласу Satellite класу Utility входять наступні екземпляри:

The screenshot shows the Protege interface. On the left, the 'Classes' browser displays a hierarchy: **rdfs:Resource** (413) > **owl:Thing** (148) > **Utility** (9) > **Satellite** (5). The **Satellite** class is selected. On the right, the 'Class Form' tab is active, showing a table of instances.

[Resource]	rdf:type	rdfs:label
◆ Connection_cost	Satellite	$0.4A + 0.2(B + C) + 0.1(D + E)$
◆ Cost_of_equipment	Satellite	$0.4B + 0.2(A + C) + 0.1(D + E)$
◆ Fee_when_purchasing_equi...	Satellite	$0.4C + 0.2(D + E) + 0.1(A + B)$
◆ Receiving_speed	Satellite	$0.4D + 0.3E + 0.1(A + B + C)$
◆ Transmission_speed	Satellite	$0.4E + 0.3D + 0.1(A + B + C)$

Рис. 17. Екземпляри класу Satellite

Клас **Wireless_Internet_Service_Provider** визначає провайдерів бездротового Інтернет, які є його екземплярами та вказує на їх кількість:

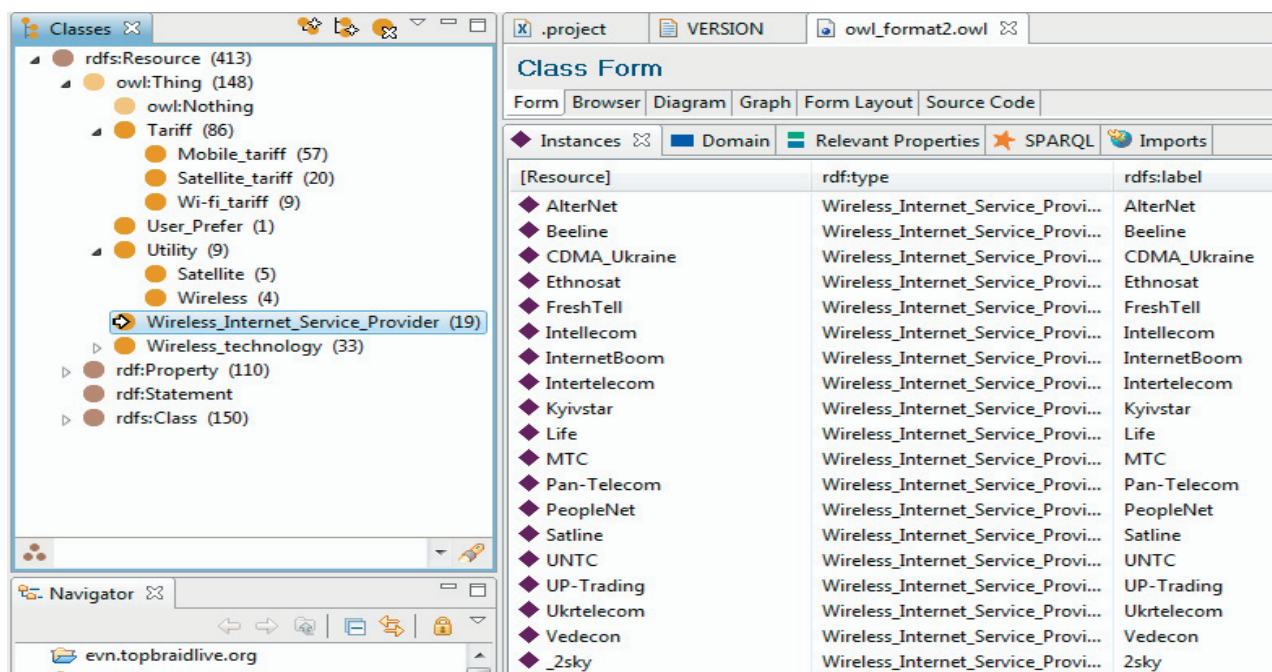


Рис. 18. Екземпляри класу Wireless_Internet_Service_Provider

Для кожного тарифного плану вказані характеристики, а саме послуги та їх вартість:

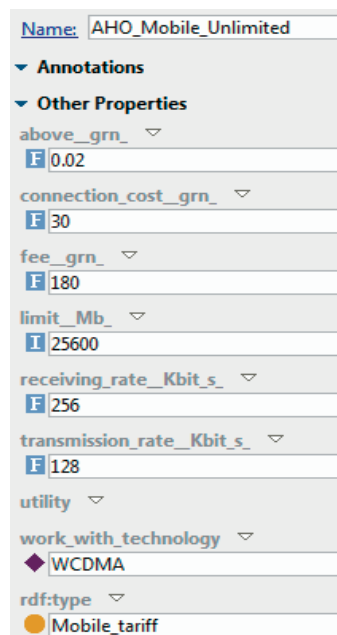


Рис. 19. Приклад характеристик тарифного плану Fresh4GB

Щоб визначати як ж саме критерії вигідності обирає для себе користувач, був створений клас User_Prefer (вподобання користувача), якому належить єдиний екземпляр current.

Даний екземпляр має наступні властивості.

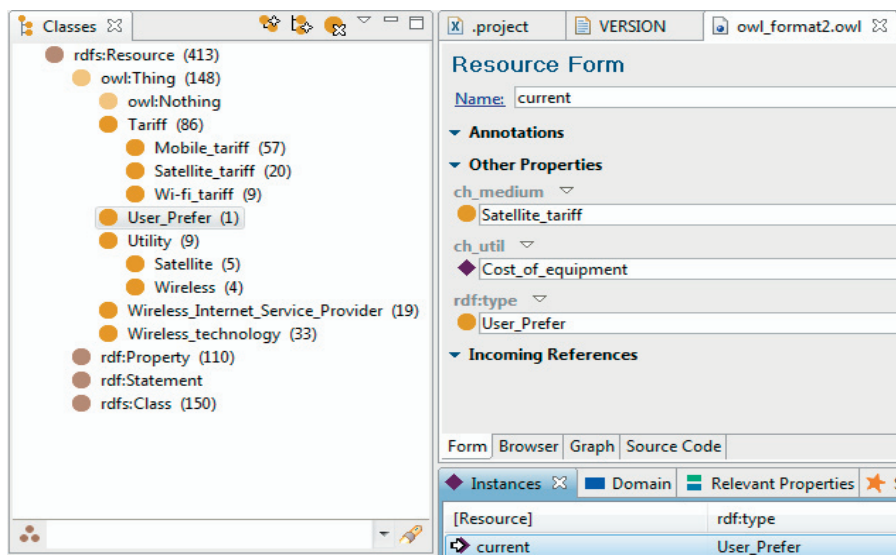


Рис. 20. Властивості екземпляру current

4. Практичні застосування результатів досліджень

Головними характеристиками інформаційно-пошукових систем є повнота та релевантність результатів пошуку.

Повнота пошуку тісно пов'язана з оперативністю охоплення інформації системою. Створена одного разу БД Інтернет-ресурсів є «зліпком» стану мережі в конкретний момент. Якщо ця база не буде оновлюватися постійно і оперативно, присутні в ній посилання на документи стануть мертвими.

Крім того, відсутність оперативності, оновлення БД не дозволить користувачеві відстежувати останні зміни у його Про.

Крім характеристик повноти та релевантності для користувачів інформаційно-пошукових систем, велике значення мають такі характеристики, як швидкість обробки запитів, отримання відгуку від системи, достовірність відгуку, а також додаткові сервіси – можливість знаходження документів, подібних вже наявним, можливість підключення автоматичних перекладів і, звичайно ж, можливість уточнення запиту безпосередньо після виконання процедури пошуку. Персоналізація інтерфейсу користувачів, що працюють в режимі он-лайн, тобто збереження їх постійних запитів і організація передплати на них, може бути реалізована на основі сучасної технології RSS (Really Simple Syndication), формату даних і технічного стандарту, який забезпечує інтегрований доступ до інформації на web-сайтах. Користувачі можуть отримати доступ до даних у форматі RSS за допомогою спеціальних програм, що називаються RSS-агрегаторами.

Для отримання тематичної стрічки (RSS-фіда) від системи InfoStream у відповідне поле RSS-агрегатора треба ввести адресу у форматі: `http://online.infostream.ua/rss.php[?<ЗАПИТ>]`, де запитом може бути слово або словосполучення на мові запитів InfoReS [1,2].

В даній роботі був вибраний RSS-агрегатор FeedReader.

Це безкоштовна програма для Windows, що дозволяє читати дані у форматі RSS версій 0.9, 0.91, 1.0, а також різну інформацію від таких систем, як Dublin Core и Slashback (стандарти опису метаданих інформаційних ресурсів мережі). На рис. 21 наведено приклад введення адреси нової RSS-стрічки:

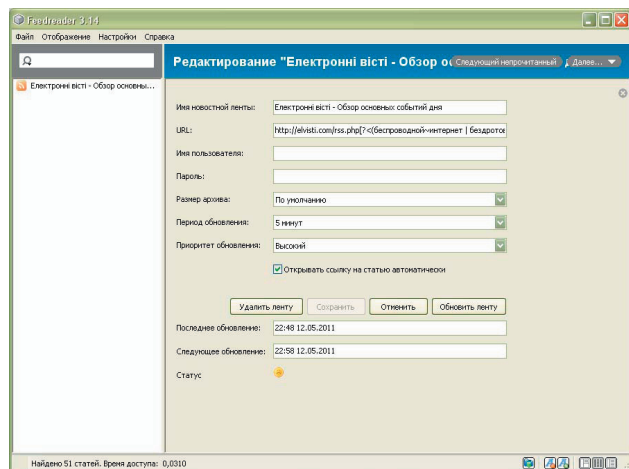


Рис. 21. Стрічка новин від системи InfoStream

На рис. 22 наведено приклад введення запиту, результат якого відповідав би тарифам та відповідним технологіям, які надають провайдери бездротового Інтернет, а отримана інформація служила б інформацією, за допомогою якої буде оновлена онтологія: (беспроводной-интернет | бездротовий-интернет) & (тарифы | тарифы).

Результатом запиту є перелік статей, в яких узані новини на задану тему.

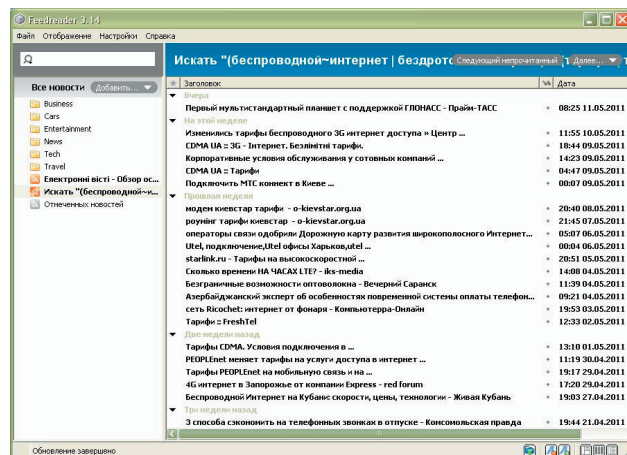


Рис. 22. Результат роботи RSS-агрегатора

Схема роботи системи вибору тарифу бездротового Інтернет наведена на рис.23:

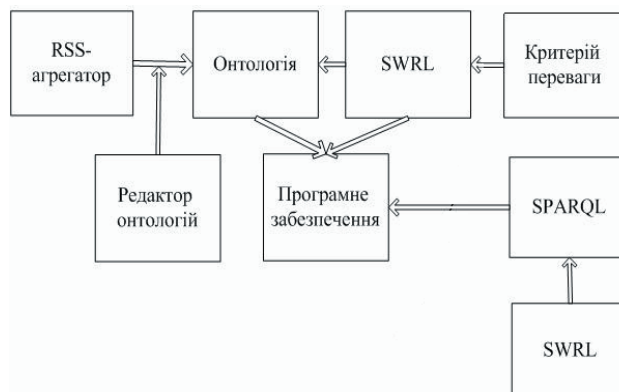


Рис. 23. Схема роботи системи вибору тарифу бездротового Інтернет за вказаним критерієм переваги

Роботу даної системи можна умовно поділити на чотири етапи:

- 1) RSS-агрегатор – онтологія.
- 2) Онтологія – програмне забезпечення.
- 3) SWRL – програмне забезпечення.
- 4) SPARQL – програмне забезпечення.

1. На першому етапі, отримавши в результаті запиту інформацію про бездротовий Інтернет в Україні, створена онтологія у разі необхідності оновлюється або доповнюється новими даними у середовищі розробки онтологій. В даній роботі для створення та редагування онтології було вибране середовище TopBraid. Це пояснюється тим, що дане середовище призначене для створення і управління моделями Про за допомогою таких стандартів Semantic Web, як RDF, RDFS та OWL. Composer – це редактор онтологій та база знань, що забезпечує візуальне редагування, а також сумісність з UML, XML Schema та базами даних. TopBraid побудований на платформі Eclipse та має Jena API. Composer об'єднує логічні, а також засновані на правилах, двигуни міркувань. Він пропонує зручний інтерфейс користувача, оснований на формах, з можливістю переглядати і редагувати онтології в різних форматах її реалізації. Перевірка, тестування і налагодження доступні завдяки вбудованим машині логічного виводу OWL, машині запитів SPARQL і машині правил. Тор-

Braid Composer дає можливість використовувати стандарти Semantic Web, імпортуючи такі моделі як XML-схема, UML, RDB-схема та таблиці. Доступні відкриті API, тому програма може працювати із серверної бази даних для покращення масштабованості.

2. Другий етап – онтологія – програмне забезпечення, на якому створена онтологія зможе підключатися до веб-інтерфейсу.

3. SWRL-правила та їх взаємодія з веб-інтерфейсом за допомогою програмного забезпечення. На даному етапі використовується клас Utility, що вказує на вигідність за тими чи іншими властивостями. Користувач зможе вибирати спочатку клас тарифів щодо їх порівняння (Wireless або Satellite). Клас Wireless містить у собі чотири екземпляри, які і будуть критеріями, за якими будуть порівнюватися тарифи між собою. Це будуть: абонплата, трафік, швидкість завантаження та швидкість передачі. Для класу Satellite екземплярами будуть вартість підключення, вартість устаткування, абон.плата, швидкість завантаження та швидкість передачі. Відповідно до цих критеріїв були створені Jena-правила, які будуть загальними для тієї чи іншої властивості, але будуть різнитися лише за класом тарифів.

Для того, щоб розрахувати вигідність тарифу, треба задати формули розрахунку.

Для класу Wireless установимо ідентифікатори властивостей для простоти використання в формулах:

- A – fee_grn
- B – limit_Mb
- C – receiving_rate_Kbit_s
- D – transmission_rate_Kbit_s

Розрахуємо коефіцієнт відповідності для кожної властивості відповідно до обраного критерію переваги:

- Абонентська плата (fee)

$$A \cdot 0,5 + B \cdot 0,003 + C \cdot 0,2 + D \cdot 0,1 \quad (1)$$

- Трафік (limit)

$$B \cdot 0,005 + A \cdot 0,3 + C \cdot 0,2 + D \cdot 0,1 \quad (2)$$

- Швидкість завантаження (receiving_rate)

$$C \cdot 0,5 + D \cdot 0,3 + A \cdot 0,2 + B \cdot 0,001 \quad (3)$$

- Швидкість передачі (transmission_rate)

$$D \cdot 0,5 + C \cdot 0,3 + A \cdot 0,2 + B \cdot 0,01 \quad (4)$$

Для класу Satellite установимо ідентифікатори властивостей для простоти використання в формулах:

- A – connection_cost_grn
- B – cost_of_equipment
- C – fee_when_purchasing_equipment_Kbit_s
- D – receiving_rate_Kbit_s
- E – transmission_rate_Kbit_s

Розрахуємо коефіцієнт для кожної властивості відповідно до обраного критерію переваги:

- Вартість підключення (connection_cost)

$$A \cdot 0,4 + B \cdot 0,2 + C \cdot 0,2 + D \cdot 0,1 + E \cdot 0,1 \quad (5)$$

- Вартість устаткування (cost_of_equipment)

$$B \cdot 0,4 + A \cdot 0,2 + C \cdot 0,2 + D \cdot 0,1 + E \cdot 0,1 \quad (6)$$

- Абонентська плата (fee_when_purchasing_equipment)

$$C \cdot 0,4 + D \cdot 0,2 + E \cdot 0,2 + A \cdot 0,1 + B \cdot 0,1 \quad (7)$$

- Швидкість завантаження (receiving_rate)

$$D \cdot 0,4 + E \cdot 0,3 + A \cdot 0,1 + B \cdot 0,1 + C \cdot 0,1 \quad (8)$$

- Швидкість передачі (transmission_rate)

$$E \cdot 0,4 + D \cdot 0,3 + A \cdot 0,1 + B \cdot 0,1 + C \cdot 0,1 \quad (9)$$

Для розрахунку критеріїв переваги були написані Jena-правила, в яких враховані коефіцієнти відповідності для кожної властивості згідно введених формул (1) – (9). Для класу Wireless правило виглядає наступним чином:

```

j0rule
[util:Wireless:(current:ch_medium?m)(current:ch_util?u)(?rdtype?m)(?fee_grn?A)(?limit_Mb?B)(?receiving_rate_Kbit_s?C)(?transmission_rate_Kbit_s?D)(?u:A?cA)(?u:B?cB)(?u:C?cC)(?u:D?cD)product(?A?cA?AA)product(?B?cB?BB)product(?C?cC?CC)product(?D?cD?DD)sum(?AA?BB?AB)sum(?CC?DD?CD)sum(?AB?CD?x)->(?utility?x)]

```

Рис. 24. Правило для розрахунку критерію переваги для класу Wireless

Перевіримо виконання правила для даної онтології. Результат має наступний вигляд:

[Subject]	Predicate	Object
AHO_Mobile_Easy	utility	154.84800332784653
AHO_Mobile_Standard	utility	175.04000389575958
AHO_Mobile_Unlimited	utility	202.40000480413437
Active_spring	utility	1288.4000410735607
Available_NL	utility	2632.400042563677
Drive	utility	1623.240024521947
Easy_start	utility	150.50000321911648
Economic	utility	161.84800343215466
Free_access	utility	2535.0000387523323
Fresh4Gb	utility	2444.096036553383
FreshDay	utility	2562.4000415205956
FreshMax	utility	2557.4000414460897
FreshOne	utility	932.4000172317028
Home1	utility	554.4000171422958
Interne-100	utility	289.7000062339939
Internet-1000	utility	304.40000654105097
Internet-500	utility	296.9000063803978
Internet_NL	utility	321.64000703394413
L	utility	409.84000834822655

Рис. 25. Результат виконання правила для класу Wireless

Правило для класу тарифів Satellite має такий же вигляд, як і для класу Wireless, але додамо ще одну властивість, що позначається як E:

```

j0rule
[util:Satellite:(current:ch_medium?m)(current:ch_util?u)(?rdtype?m)(?connection_cost_grn?A)(?cost_of_equipment_grn?B)(?fee_when_purchasing_equipment_grn?C)(?receiving_rate_Kbit_s?D)(?transmission_rate_Kbit_s?E)(?u:A?cA)(?u:B?cB)(?u:C?cC)(?u:D?cD)(?u:E?cE)product(?A?cA?AA)product(?B?cB?BB)product(?C?cC?CC)product(?D?cD?DD)product(?E?cE?EE)sum(?AA?BB?AB)sum(?CC?DD?CD)sum(?AB?CD?AC)sum(?AC?EE?x)->(?utility?x)]

```

Рис. 26. Правило для розрахунку критерію переваги для класу Satellite

Результат виконання правила наведених на рис. 27.

[Subject]	Predicate	Object
Business	utility	5531.600082427263
Business_Standard_128	utility	4469.880071489215
Corporate_1024	utility	5872.560048445463
Corporate_128	utility	4633.260078806579
Corporate_512	utility	5434.440120041967
Corporate_Four	utility	5833.000086918473
Corporate_One	utility	6661.800099268556
Corporate_Three	utility	5017.000074759126
Corporate_Two	utility	4609.000068679452
Five	utility	4800.800071537495
Four	utility	4502.400067090988
Home	utility	2189.0200399431587
Home_Standard_128	utility	4413.780064549745
Home_Unlimited	utility	4495.380071869195
One	utility	4026.000059992075
Six	utility	5257.200078338385
Sputnik	utility	7526.760109716058
Three	utility	4242.200063213706
Two	utility	4064.80006057024

Рис. 27. Результат виконання правила для класу Satellite

4. Для того, щоб вивести користувачу результат порівняння, використовується мова запитів SPARQL. Виводитись будуть тільки перші п'ять тарифів, що мають найбільший чи найменший коефіцієнт, розрахований за формулами, в залежності від обраної властивості. Синтаксис запиту виглядає наступним чином:

```

SELECT ?tariff ?provider ?utility
WHERE {
  :current :ch_medium ?ch .
  ?tariff rdfs:type ?ch ;
  :utility ?utility .
  ?provider rdfs:type :Wireless_Internet_Service_Provider ;
  :has_tariff ?tariff
} ORDER BY DESC (?utility) LIMIT 5

```

Рис. 28. Синтаксис запиту

Наведемо приклад роботи системи вибору найбільш вигідного тарифу провайдера бездротового Інтернет відповідно до обраного користувачем критерію переваги. Установимо значення ch_medium як Satellite_tariff, a ch_util як Cost_of_equipment:

tariff	provider	utility
Home	Intelcom	2189.0200399431...
One	_2sky	4026.0000599920...
Two	_2sky	4064.80006057024
Three	_2sky	4242.2000632137...
Home_Standard_...	Ethnosat	4413.7800645497...

Рис. 29. Результат виконання запиту

Установимо значення ch_medium як Mobile_tariff, a ch_util як Fee:

tariff	provider	utility
Easy_start	PeopleNet	88.600000971928...
MTC_Connect-50	MTC	97.600001070648...
_3G_600	Kyivstar	98.800001088529...
AHO_Mobile_Easy	Ukrtelecom	100.14400100708...
Economic	PeopleNet	117.64400100708...

Рис. 30. Результат виконання запиту

Установимо значення ch_medium як Wi-fi_tariff, a ch_util як Receiving_rate:

tariff	provider	utility
_1	Vedecon	3420.0000542998...
AHO_Wifi_10Gb	Ukrtelecom	522.00001111626...
_0410	Beeline	169.47500157414...
_0405	Beeline	168.04000155162...
_7755	Kyivstar	167.90500154846...

Рис. 31. Результат виконання запиту

Висновки

В результаті проведеного аналізу побудована онтологія Про «Бездротовий Інтернет в Україні», яка оновлюється за допомогою технологій контент-моніторингу та відображає актуальну інформацію. За допомогою онтологічного підходу вся інформація про бездротовий Інтернет в Україні розкласифікована відповідно до поколінь мобільного Інтернет, технологій підключення та, в кінцевому рахунку, тарифів, які пропонують наразі Інтернет-провайдери; для кожного Інтернет-провайдера було створено перелік всіх тарифів, які він пропонує, та вказані характеристики, значення яких однозначно ідентифікують кожний тариф.

Щоб користувач зміг обирати тарифи за вказаним ним критерієм переваги та класом тарифів, використовуючи формули для розрахунку вигідності тарифу, було створено SWRL-правила та Jena-правила, котрі включають в себе властивості, які може обрати користувач.

Використовуючи SPARQL-запит, були порівняні та виведені тарифи, які найбільше відповідають запиту користувача.

Результати виводяться за допомогою web-інтерфейсу.

Розроблено узагальнений алгоритм автоматичного вибору найбільш відповідного тарифу на основі онтологічного підходу за допомогою технології контент-моніторингу, що включає наступні кроки: побудова загальної онтології, в якій представлені всі існуючі технології бездротового Інтернет, Інтернет-провайдери, послуги та їх характеристики; відстеження та виявлення в Інтернет-джерелах всіх оновлень інформації про послуги і тарифи Інтернет-провайдерів; оновлення існуючої онтології; створення swrl-правил; порівняння тарифів на основі вибраних параметрів за допомогою swrl-правил; визначення найбільш вигідної пропозиції на ринку бездротового Інтернет; виведення результату за допомогою SPARQL-запиту.

Отже, контент-моніторинг дає змогу експертам оновлювати та поповнювати онтологію, а середовище конструювання онтології – її редагувати. За допомогою бібліотеки Jena можна створювати правила, основані на формулах або на певних умовах, а результат виводити за допомогою web-інтерфейсу, для того, щоб користувач міг бачити результати роботи подібних систем.

Література

1. Ландэ Д.В. Основы интеграции информационных потоков [Текст] / Д.В. Ландэ. - К.: Инжиниринг, 2006. - 240 с.
2. Григорьев А.Н. InfoStream. Мониторинг новостей из Интернет: технология, система, сервис [Текст]: науч.-метод. пос. / А.Н. Григорьев, Д.В. Ландэ, С.А. Бороденков, Р.В. Мазуркевич, В.Н. Пацьора – К.:ООО «Старт-98», 2007. – 40 с.
3. Рябова Н.В. Методы и модели интеллектуальной обработки текстов в задачах онтологического инжиниринга [Текст] / Н.В. Рябова // Матеріали YI Міжн. наук.-практ. конф. «Математичне та програмне забезпечення інтелектуальних систем». – Дніпропетровськ: ДНУ. – 2008. - с.269-270.
4. Рябова Н.В. Методы и модели интеллектуального анализа текстов в задачах онтологического инжиниринга [Текст] / Н.В. Рябова, В.В. Волкова, Я.В. Дыдыкина // Материалы II Межд. радиоэлектронного Форума «Прикладная радиоэлектроника. Состояние и перспективы развития», МРФ–2005. – Харьков, Украина: в 3-х т. – Т.3. – 2005. – с.87-90.
5. Рябова Н.В. Применение интеллектуальных информационных технологий в задачах выбора провайдера беспроводного Интернет [Текст] / Н.В. Рябова, Ю.В. Сидоренко // YI Межд. научн.-практ. конф. «Наука и социальные проблемы общества: информатизация и информационные технологии». Сб. научн. трудов – Х.: ХНУРЭ, 2011. – с.134-135.

Abstract

The article concerns the use of content monitoring technology of Internet resources for ontology clarifying and updating of given Pro. The ontology “Wireless Internet in Ukraine”, which is constantly updating and is used to solve the problem of choosing of the optimal tariff and internet service provider (ISP), was developed. A generic algorithm of the problem solution was designed. It includes following steps: designing of a general ontology, which represents all existing technologies of wireless Internet; Internet service providers; services and their characteristics; tracing and identification of information updating about services and tariffs of Internet service providers; updating of existing ontology; SWRL-rules making; comparison of tariffs on the basis of parameters selected with SWRL-rules and detection of the most advantageous offer for the wireless Internet market. Test examples of the results of the algorithm application, using SWRL-rules and SPARQL-queries, are given

Keywords: ontology, content monitoring, wireless Internet, SWRL-rule, SPARQL-query