

НПК МНІС ІП-2018

ЗБІРНИК НАУКОВИХ ПРАЦЬ
МОЛОДИХ
НАУКОВЦІВ І СТУДЕНТІВ

ЧАСТИНА

1

ПРИСВЯЧУЄТЬСЯ 30-РІЧЧЮ
ПІДГОТОВКИ ІТ-ФАХІВЦІВ В
ЖМЕЛЬНИЦЬКОМУ
НАЦІОНАЛЬНОМУ
УНІВЕРСИТЕТІ



МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Хмельницький національний університет

Військовий інститут Київського національного університету
ім.Тараса Шевченка

ПВНЗ “Університет економіки і підприємництва”

Тернопільський інститут агропромислового виробництва

Інтелектуальний потенціал - 2018

збірник наукових праць молодих науковців і студентів

Присвячується 30-річчю підготовки ІТ- фахівців в Хмельницькому національному університеті

сформовано за матеріалами

Всеукраїнської науково-практичної конференції
молодих науковців і студентів «Інтелектуальний потенціал – 2018»

14-16 листопада 2018р.

Частина 1

Комп’ютерні науки та інформаційні технології проектування

Хмельницький
2018

ББК 74.480.278

С.88

«Інтелектуальний потенціал – 2018» - збірник наукових праць молодих науковців і студентів з нагоди 30-річчя підготовки ІТ-фахівців в ХНУ/Колектив авторів – Хмельницький: ПВНЗ УЕП, 2018. – Ч.1: Комп'ютерні науки та інформаційні технології проектування. – 132 с.

Відповідальний редактор: Капітанець С.В.

Відповідальний за випуск: Чещун В.М.

Редакційна колегія:

Желавський О.Б.

Капітанець С.В.

Мясіщев О.А.

Чещун В.М.

Тімофєєва Л.В.

ЗМІСТ

Бакаляр А.В., Манзюк Е.А., Скрипник Т.К. Сучасні методи використання комп'ютерного зору	5
Башта А.Р., Свірневський М.С. Структура клієнт-серверної системи забезпечення взаємодії мобільних пристроїв	8
Березніцький С.О., Бармак О.В. Інформаційна система для візуалізації смислової складової текстового контексту	11
Білий Д.І. Система автоматизованого проектування CAD/CAE додатків	15
Болкун А.О., Міхалевський В.Ц. Автоматизована система для закладу з надання косметологічних послуг	18
Гашук Т.О., Бармак О.В. Інформаційна технологія визначення ознак обличчя, що впливають на прояви емоцій	22
Гикавчук А.В., Свірневський М.С. Система підтримки та аналізу інформаційних потоків суспільної організації	28
Дем'янюк М.В., Багрій Р.О. Інформаційна система миттєвого обміну повідомленнями в локальній мережі	31
Джурабаєв О.В., Бармак О.В., Манзюк Е.А. Пошук змісту в текстовій інформації	35
Колошинський Д.В., Багрій Р.О. Інформаційна система для роботи з сервісами хмарних сховищ	39
Кухарчук М.Н, Пасічник О.А., Скрипник Т.К. Експертна система закладів громадського харчування	43
Лесишин М.В., Манзюк Е.А., Скрипник Т.К. Інформаційна технологія класифікації зображень на базі SVM	48
Мазурець О.В., Ковальчук О.В., Слободзян В.О., Білоус Г.А. Метод формального опису елементів моделей автоматизованого формування тестових завдань	51
Мартинюк Б.І., Лищук О.А. Система підбору методом спільної фільтрації продукції інтернет-магазину літератури	56
Медцовський О. В. Інформаційна система забезпечення контролю фінансів ведення малого бізнесу	60
Мельник М.О., Плетюк М.М., Коломієць Н.О. Інформаційна система генерування автоматизованих рекомендацій на базі активності користувача	63
Мосьондз В.О., Петровський С.С. Особливості впровадження сучасних експертних систем	66

Ніколайчук В.А., Мазурець О.В. Адаптивний метод аналізу кольорових зображень при нейромережевому розпізнаванні зашумлених образів	69
Пшедзьял Я.Е., Скрипник Т.К. Інформаційна система забезпечення ефективності функціонування бізнес-процесів організації на прикладі кінотеатру.....	74
П'ятін В.Д., Свірневський М.С. Інформаційна система автоматизованого реферування	80
Рибачук В.В. Мережеві технології забезпечення відображення віртуального ігрового контенту.....	83
Рижак А.В., Лищук О.А. Використання методу ансамблів для класифікації.....	85
Ройзнер К.І., Мазурець О.В. Використання методу локалізації при автоматизації діяльності релігійної установи.....	89
Сергієва О.О., Мазурець О.В. Інформаційна технологія рекурсивного семантичного аналізу текстів шляхом дисперсійного оцінювання слів..	93
Ставнійчук М.В., Мазурець О.В. Інформаційна модель автоматизації супроводу навчального процесу.....	98
Терлецький Ю.В, Свірневський М.С., Скрипник Т.К. Інформаційна технологія класифікації текстової інформації.....	103
Тимчак А.В., Лищук О.А., Скрипник Т.К. Інформаційна технологія автоматизованого випадково генерування корпусів відповідно до гаусівського розподілу	107
Торчинський О.І., Пасічник О.А. Інформаційна технологія формування компетентностей та прогнозування тенденцій вимог до фахівців ІТ галузі.....	110
Фарина А.П., Мясіщев О.А. Аналіз методу тадеуса вінсенті для вирішення прямої геодезичної задачі на еліпсоїді.....	114
Шкута В.О., Багрій Р.О., Скрипник Т.К. Інформаційна технологія для збільшення кількості переглядів оголошень з нерухомості	116
Шлапак О.В., Манзюк Е.А. Особливості використання методу випадкового лісу для визначення шаблонів в потоці даних	120
Шоханов А.С., Міхалевський В.Ц. Автоматизація побудови гібридної інфраструктури бази Microsoft Exchange Server та Office 365.....	123
Яцунь В.В. Інформаційна технологія для формування автоматизованих корпусів з мережі Інтернет	128

Сучасні методи використання комп'ютерного зору

Бакаляр А.В., Манзюк Е.А., Скрипник Т.К.

Хмельницький національний університет

Комп'ютерний зір (computer vision, машинний зір, технічний зір, зір роботів) оформилося як самостійна дисципліна до кінця 60х років. Цей напрямок виник у рамках штучного інтелекту в той його період, коли ще були гарячі суперечки про можливість створення мислячої машини. Воно виділилося з робіт з розпізнавання образів, коли була усвідомлена його математична специфіка, фотозображення являє собою двомірну проекцію тривимірного миру, тобто виродженого перетворення. Це означає, що для повної та правильної інтерпретації того, що зображене, необхідно мати додаткову інформацію. За аналогією із зором людини, у якості способів одержання додаткової інформації стали розглядати рух зорового обладнання (активний огляд), використання стереопари, апріорних моделей побаченого як геометричних, так і змістовних, семантичних. Цей підхід в усьому світі був штучним, тому що ні техніки, ні математичного забезпечення для розв'язку таких складних завдань ще не було. Поряд із цим джерелом, машинний зір мав і інші проблеми, технічні та практичні. Фотозображення почали використовувати для вимірів задовго до винаходу комп'ютера (фотограмметрія). Цей підхід використовувався для аерофотознімання, пізніше – для космічної навігації та інших завдань. При розв'язку практичних завдань використовувалися трохи інші способи одержання додаткової інформації. Це в різних видах штучна організація середовища, у тому числі спеціальне підсвічування (наприклад, щілинне підсвічування для полегшення виділення ребер), використання різного роду маркерів і маяків (наприклад, триангуляційні вишки при аерофотозніманні або маркування деталей на конвеєрі), стандартне розташування предметів.

Ціль роботи дослідження методів та підходів компютерного зору.

Дуже багато сучасних авторів у своїх роботах ґрунтуються на результатах, отриманих або М. Поллефейсом, або Д. Ністером. Алгоритм, запропонований першим автором, дає відмінну якість на відеопослідовностях, отриманих з довільних камер. Тобто про внутрішні параметри камер на момент початку відновлення немає ніякої інформації. Алгоритм, запропонований другим автором, працює з відкаліброваними камерами та дає відновлення сцени в реальному часі [2].

Методи різняться у виборі алгоритмів. Для першого:

- виділення та співставлення особливостей – KLT алгоритм;
- для пошуку фундаментальної матриці використовується RANSAC;
- для генерації гіпотез використовується 8-точковий алгоритм;
- для уточнення матриці, отриманої на виході Ransac'a, використовується метод Левенберга-Макварда;
- для одержання тривимірних координат точок використовується

триангуляція, запропонована Хартлі та Штурмом, і заснована на мінімізації відстані від проєкцій точок на зображенні до відповідних епіполярних ліній.

Ністер використовує наступні методи:

- виділення та співставлення особливостей – детектор Харріса;
- для пошуку фундаментальної матриці на підставі відповідностей використовується алгоритм запропонований Д.Ністером для можливості обробки в реальному часі;
- для генерації гіпотез використовується 8-точковий алгоритм;
- уточнення на виході не проводиться;
- для одержання тривимірних координат точок використовується триангуляція Ністера.

У даній роботі за основу брався метод Поллефейса, з деякими змінами. Тут відсутній етап самокалібрування та злиття результатів, отриманих для різних пар зображень.

Лінійна триангуляція.

Нехай координати проєкції точки на перше зображення $x_1 = PX = w(u, v, 1)$. Якщо позначити за p_i^T i -й рядок матриці P , рівняння $x_1 = PX$ можна переписати у вигляді:

$$wu = p_1^T X, wv = p_2^T X, w = p_3 X$$

Забираючи w з перших двох рівнянь, одержуємо два рівняння:

$$up_3^T X = p_1 X$$

$$vp_3^T X = p_2$$

Аналогічні два рівняння можна одержати для проєкції точки на друге рівняння. Таким чином, можна записати систему $AX=0$, де A – матриця 4×4 .

Скоріше всього, через шум, точний розв'язок, використовуючи цю систему, одержати результат не вдасться.

Існує два способи пошуку розв'язків системи. Перший – вирішувати систему, використовуючи МНК для випадку однорідної системи. Другий – вважаючись, що $X = (x, y, z, 1)$, одержуємо неоднорідну систему на x, y, z яка також вирішується МНК для неоднорідної системи.

Мінімізація суми квадратів відстаней до епіполярних ліній.

Нехай e відповідні точки u і u' . Кожне зображення перетвориться так, щоб $u = u' = (0, 0, 1)$, а епіполюси мали координати $e = (1, 0, f)$ та $e' = (1, 0, f')$.

Якщо T та T' – ці перетворення, між новими зображеннями буде $F' = T'FT^{-1}$. Вводиться параметризація епіполярних ліній. Епіполярна лінія $l(t)$ – лінія минаюча через $(0, t, 1)$ і $(1, 0, f)$ має вигляд:

$$l(t) = e \times u = (tf, 1, -t)$$

Знаходимо координати відповідної лінії на другому зображенні, виражаємо відстань до початку координат і одержуємо величину виду:

$$s(t) = \frac{t^2}{1 + (tf)^2} + \frac{(ct + d)^2}{(at + b)^2 + f'^2 (ct + d)^2} \quad (1)$$

Далі береться похідна $r(t)$ і прівірюється до 0:

$$r(t) = t((at + b)^2 + f'^2 (ct + d)^2)^2 - (ad - bc)(1 + (tf)^2)^2 (at + b)(ct + d) \quad (2)$$

Одержуємо рівняння 6-го ступені на t . Один з її нулів відповідає мінімальному значенню $s(t)$.

У цьому проекті в основному використовували OpenCV для попередньої обробки даних зображення, наприклад, читати зображення в масив і змінювати розмір, який нам потрібен.

Звичайним способом поліпшення результатів навчання зображень є деформація, обрізка або освітлення навчальних матеріалів випадковими способами. Ця перевага полягає у збільшенні ефективного розміру навчальних даних завдяки всім можливим варіаціям тих самих зображень і, як правило, допомагає мережі навчитися впоратися з усіма спотвореннями, які відбудуться в реальному використанні класифікатора.

З результату можемо побачити:

– У k -NN чіткість точок з пікселями та точність гістограми порівняно однакові. На 5 підкласах під назвою мітки точність гістограми трохи вища, ніж сирий піксель, але над усім, сирий піксель показує кращий результат.

– У класифікаторі MLP нейронної мережі сигнал точності пікселя значно нижчий, ніж точність гістограми. Для всього набору даних точність порожніх пікселів навіть нижча, ніж випадкові здогади.

– Усі ці два методи `sklearn` не дають дуже хорошої продуктивності, точність визначення правильної категорії становить лише 24% усього набору даних. Ці результати показують, що використання методів `sklearn` для розпізнавання зображень є недостатнім. Вони не здатні забезпечити хорошу продуктивність для складних зображень з багатьма категоріями. Але, порівнюючи їх із випадковими догадками, вони досягли певного поліпшення.

Література

1. Шапиро Л. Компьютерное зрение / Л. Шапиро, Дж. Стокман; пер. с англ. – М.: Бином. 2006. – 752 с.
2. Гороховатский В.А. Распознавание изображений в условиях неполной информации / В.А. Гороховатский. – Харьков: ХНУРЭ, 2003. – 112 с.
4. Путьгин Е.П. Обработка изображений в робототехнике / Е.П. Путьгин, С.И. Аверин. – М.: Машиностроение, 1990. – 320 с.

Структура клієнт-серверної системи забезпечення взаємодії мобільних пристроїв

Башта А.Р., Свірневський М.С.

Хмельницький національний університет

Розробка клієнт-серверної системи забезпечення взаємодії мобільних пристроїв та її використання є актуальним питанням на сьогоднішній день, оскільки мільйони людей щодня, не виходячи з дому, купують різні товари в електронних магазинах.

Метою роботи є аналіз специфіки діяльності мобільних клієнт-серверних систем, їх відмінності від інших способів онлайн-торгівлі, розглянути сучасні засоби по розробці мобільних додатків та розробити клієнт-серверну систему забезпечення взаємодії мобільних пристроїв.

Новизною роботи є застосування сучасних мобільних технологій, а саме мови програмування Swift для розробки мобільного додатку для iOS та систему управління базами даних MySQL для розробки бази даних магазину.

В світі загалом та зокрема в Україні величезними темпами росте кількість користувачів мобільних пристроїв і, як наслідок, кількість «електронних» покупців. Це обумовлено перевагами, що відображені на рисунку 1.



Рисунок 1 – Рейтинг переваг клієнт-серверних систем.

Клієнт-серверна система забезпечення взаємодії мобільних пристроїв – це реалізоване в мережі представництво, шляхом створення клієнта і сервера для продажу товарів і послуг іншим користувачам. Інакше кажучи, це співтовариство територіально роз'єднаних співробітників магазину (продавців, касирів) і покупців, які можуть спілкуватися і обмінюватися інформацією через електронні засоби зв'язку при повній (або мінімальній)

відсутності особистого прямого контакту.

Щоб реалізувати вказані вище переваги (див. рис.1), використовуючи можливість клієнт-серверної система забезпечення взаємодії мобільних пристроїв структурна модель такої системи повинна складатися з наступних модулів (рис.2):

- програмний модуль роботи з адміністративною частиною;
- програмний модуль клієнтської логіки;
- програмний модуль доступу до бази даних (БД).

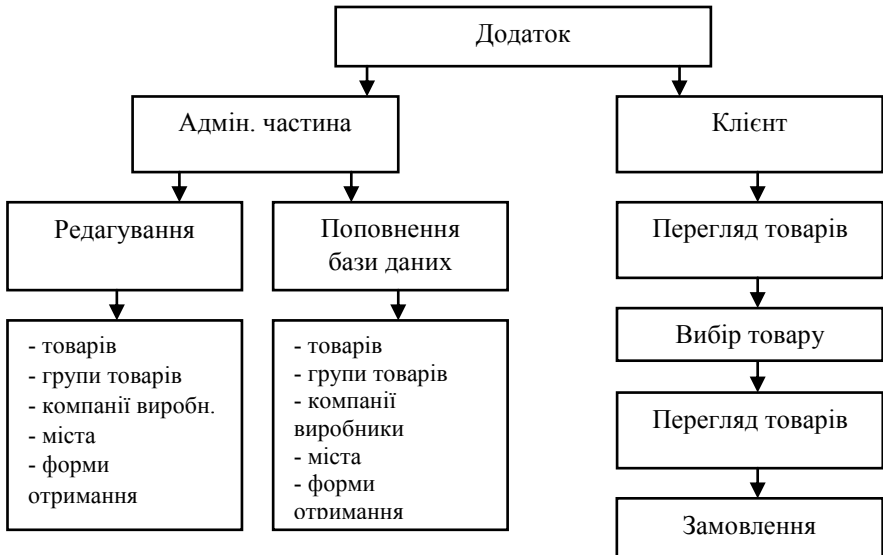


Рисунок 2 – Структурна модель клієнт-серверної системи

Програмний модуль для роботи з інтерфейсом забезпечує коректне відображення інформації продукції на сторінках та самих сторінок [1].

Програмний модуль головної логіки необхідний для коректної роботи посилань та кнопок, що дозволяють реєструвати користувача та здійснювати входження в систему із переходом на різні сторінки додатку, а саме:

- каталог товарів;
- сторінка кожного товару окремо;
- «кошик»;
- оформлення замовлення;
- контактна інформація.

Сторінки каталогу товарів є основою мобільного магазину. Складовою цих сторінок, зазвичай є фото товарів, ціна, назва, кнопка «Купити» або «Додати в кошик». На цій сторінці можна розмістити товари по рейтингу [2].

На сторінці товару повинні бути деталізовані фото товару, його назва, ціна, основні характеристики, наявність у продажу.

«Кошик» – найважливіша фінансова складова мобільного магазину. Він повинен бути легким у використанні. При бажанні закінчити замовлення користувач перенаправляється на сторінку закінчення оплати. Також має бути у наявності кнопка «очистити корзину».

При реєстрації покупця потрібно отримувати від нього лише ті дані, які необхідні для здійснення оплати та доставки. Чим простіша форма реєстрації, тим більша ймовірність того, що користувач здійснить покупку. Обов'язкові дані – ПІБ, номер телефону, адреса доставки, e-mail [3].

У додатку повинні бути й інші сторінки – сторінка відгуків, сторінка для питань покупців, сторінка переліку акцій, тощо.

Для ефективного функціонування магазину у БД ще потрібно зберігати дані про:

- товари, що доступні для продажу (таб. «Products»);
- користувачів (таб. «Users»), що бажають, або вже здійснили покупку;
- товари (таб. «Card»), які було додано в корзину певним користувачем.

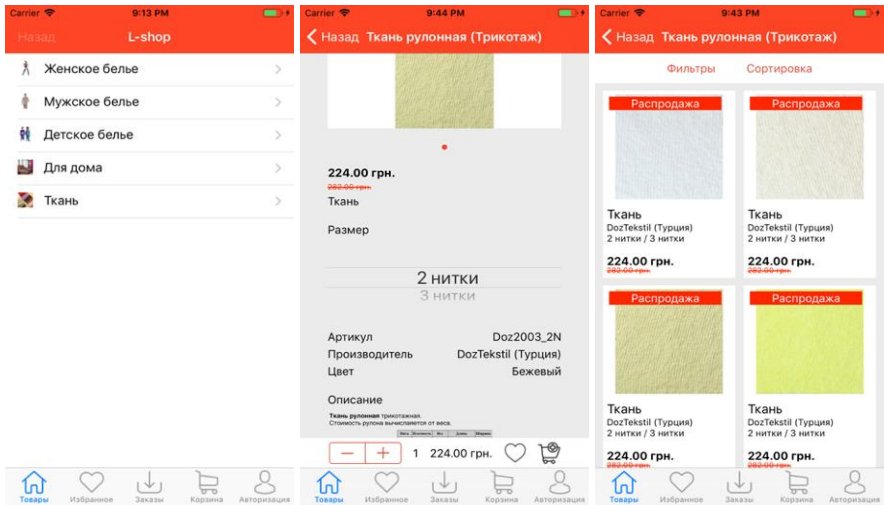


Рисунок 3 – Інтерфейсна частина програми.

Враховуючи аналіз переваг і можливостей технологій онлайн-торгівлі для створення мобільного магазину, сучасну технологію мови програмування Swift та системи управління базами даних MySQL, було розроблено програму “L-shop”, функціональна структура якого складається з мобільного

додатку та бази даних, дані елементи взаємодіють між собою. На основі проведеного дослідження була обрана оптимальна архітектура клієнт-серверної системи забезпечення взаємодії мобільних пристроїв і був розроблений додаток для електронної торгівлі за допомогою програми xCode 10.0 та мови програмування Swift 4.2.

Література

1. Горев А. Эффективная работа с СУБД. / А. Горев, С. Макашарипов, Р. Ахаян. – Питер: СПб, 2003. – 445 с. – ISBN: 5-88782-132-9.
2. Грамотная структура мобильного магазина [Электронный ресурс]. – Режим доступа: <http://prodex.ua/blog>
3. Меджибовська Н. С. Електронна комерція: Навч. посібник / Н. С. Меджибовська – К.: Центр навчальної літератури, 2004. – 384 с. – ISBN 966-85680-0-1.

Інформаційна система для візуалізації смислової складової текстового контексту

Березницький С.О., Бармак О.В.

Хмельницький національний університет

З розвитком нових технологій та постійним підвищенням рівня інформатизації суспільства проблема машинної класифікації та визначення сенсу тексту набуває особливого значення. Обробка і класифікація тексту може використовуватися в пошукових системах, на сайтах, в бібліотеках та в базах даних, де користувачу необхідна інформація буде знаходитися не по ключових словах, а за змістом самого ресурсу. Методи машинного навчання дозволяють розуміти семантичну структуру ресурсу, що дозволяє суттєво збільшити результативність пошуків. Для ефективного дослідження даної проблеми необхідно бачити результати навчання системи з різними наборами даних.

Візуальне відображення даних є набагато інформативнішим за інші методи отримання та сприйняття інформації, більш зручним і легшим для сприйняття та розуміння ніж наприклад відображення даних у таблицях, схемах, математичних матрицях, або просто в числах. У наш час поширюються засоби для візуального представлення інформації та даних і з такою метою широко використовується інформаційні технології, тому виникає потреба створювати програмне забезпечення для візуалізації даних. З'явилися також нові різновиди візуального представлення даних, деякі з них походять від графіків, діаграм, гістограм і не просто візуалізація на площині, а візуалізація у 3D просторі.

Часто задача класифікації тексту є окремим елементом складної проблеми, наприклад, створення інтелектуального консультанта чи чат-бота. В цьому випадку є багато повідомлень, розділених на декілька класів.

Метою роботи є розробка інформаційної системи для класифікації та пошуку смислової складової тексту та візуалізації отриманих результатів для подальшого аналізу, що забезпечує максимально точне передання змісту вхідного тексту.

Було запропоновано інформаційну систему для побудови системи лінійної класифікації у вигляді дерева рішень. Вхідними даними для застосування інформаційної технології для візуалізації смислової складової текстового контексту є текстів, які належать до певних категорій та будуть використані в якості навчальної вибірки, та тестовий матеріал який необхідно класифікувати.

В основі функціонування інформаційної системи візуалізації смислової складової текстового контенту є використання дерева рішень [1]. Для цього спочатку складається набір правил за якими буде працювати алгоритм дерева рішень. Для знаходження правил для дерева рішень необхідно виконати наступні кроки.

1. За допомогою методу опорних векторів [2] формується вхідна модель даних. Кожен текст з навчального набору аналізується методом опорних векторів і відповідно визначається його відсоткова належність ко кожній з категорій. В результаті кожен текст можна назвати n -мірним вектором, де відсоткова належність до кожної категорії є відповідною координатою n -мірного простору.

2. З вхідної моделі формується діагональна матриця відстаней між кожною статтею.

3. Визначення координат кожного з текстів у 2-вимірному просторі(1), для можливості сприйняття і аналізу вхідних даних (рисунок 1):

$$\{x_1 \ x_2 \dots x_n \} \rightarrow \{x_1 \ x_2 \} \quad (1)$$

Для цього було використано алгоритм багатомірного шкалювання [3], на вхід якого подається матриця матриця, елемент якої на перетині i -го рядка і j -го стовпця, містить відомості про попарну схожість аналізованих об'єктів (об'єкта [i] і об'єкта [j]), в даному це матриця відстаней між кожним текстом. На виході алгоритму ідуть числові значення координат, які приписуються кожному об'єкту в деякій новій системі координат (рисунок 1).

В результаті багатомірного шкалювання можна спостерігати що на площині утворилися групи текстів, які належать одній категорії (рисунок 2).

4. Виділення категорій за допомогою кривої з прямих ліній, які є правилами для алгоритму дерева рішень (рисунок 3).

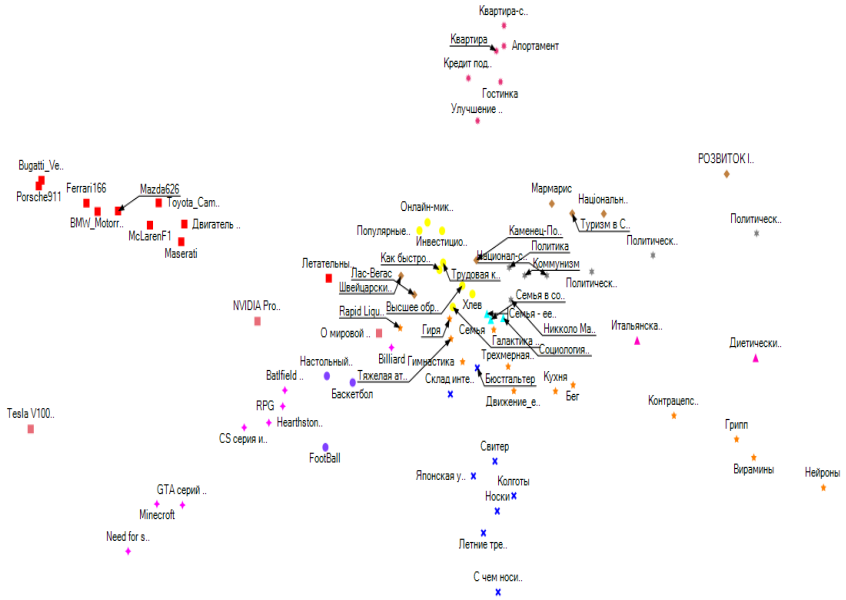


Рисунок 1 – Відображення текстів на площині

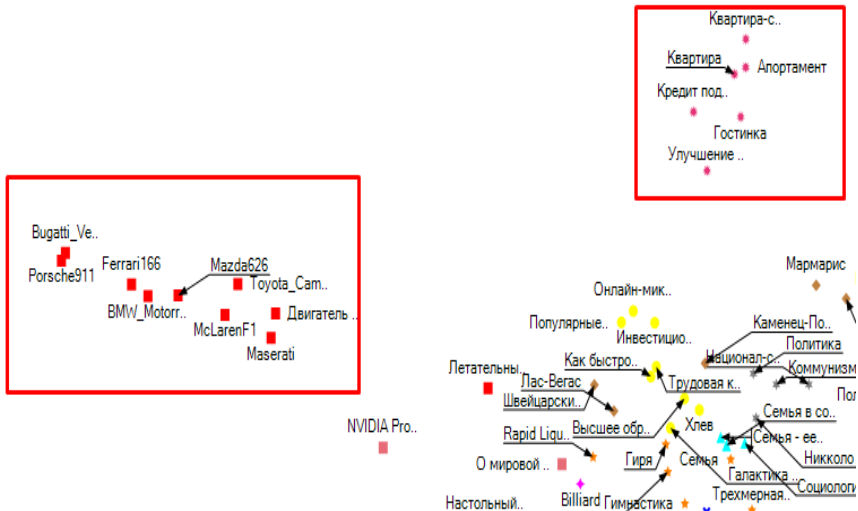


Рисунок 2 – Групи тестів які належать до певної категорії

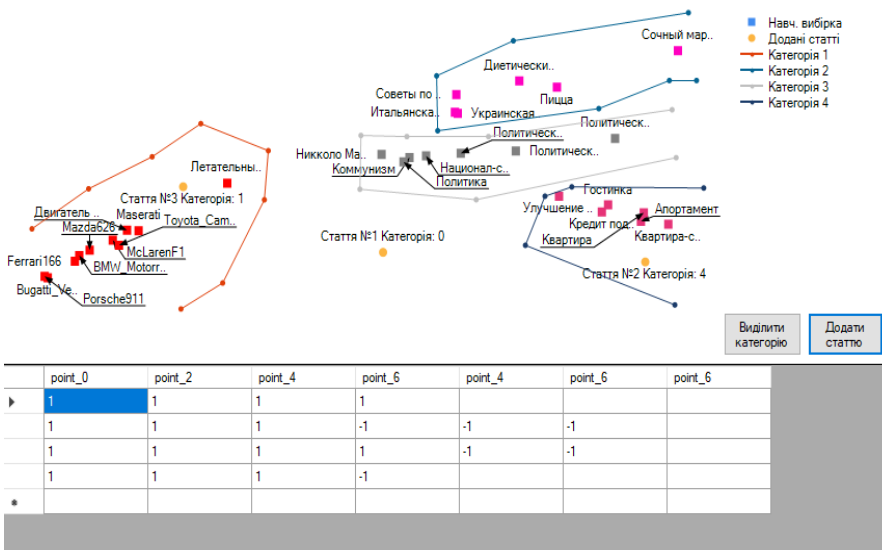


Рисунок 3 – Виділення категорій та формування правил

Відповідно до формули прямої (2) можна визначити де знаходиться точка відносно неї, відповідно, кожна категорія має набір правил, значень -1,0,1, які позначають, де повинна знаходитись точка тесту відносно кожного сегменту кривої.

$$kx + b - y = 0 \quad (2)$$

При виконанні всіх правил категорії можна стверджувати, що даний тест належить до неї.

Отже, запропонована інформаційна система для візуалізації смислової складової текстового контексту забезпечує точну і швидку класифікацію тексту відповідно до заданих категорій.

Подальші дослідження спрямовані на автоматизацію процесу виділення категорій та формування правил на площині, що забезпечить можливість інтеграції даної інформаційної системи в програмні комплекси.

Література

1. Використання "дерева рішень" [Електронний ресурс]. – Режим доступу: pidruchniki.com/10780621/ekonomika/vikoristannya_dereva_rishen
2. Метод опорных векторов [Електронний ресурс]. – Режим доступу: https://www.machinelearning.ru/wiki/index.php?title=Метод_опорных_векторов
3. Многомерное шкалирование [Електронний ресурс]. – Режим доступу: <http://statsoft.ru/home/textbook/modules/stmulasca.html>

Система автоматизованого проектування CAD/CAE додатків

Білий Д.І.

Хмельницький національний університет

Системи автоматизованого проектування стали невід'ємною складовою частиною машинобудівного підприємства. Тому створення нових та вдосконалення існуючих CAD/CAE систем є актуальною задачею, яка потребує постійної уваги розробників програмного забезпечення. Процес вдосконалення систем автоматизованого проектування продовжується постійно, тому на етапі побудови, особливо важливим є етап вибору архітектури CAD/CAE системи, який би дозволяв не змінюючи архітектуру системи масштабувати її, вносити зміни та доповнення до елементів системи.

Одним із підходів, який використовується при створенні програмного забезпечення є використання архітектурних шаблонів. Найбільш відомим архітектурним шаблоном є MVC (model-view-controller), який використовується при написанні серверних додатків [1].

Для створення CAD/CAE систем існує архітектурний шаблон MVO (model-view-operator) [2].

Метою роботи є аналіз архітектурного шаблону проектування MVO з метою його використання для створення додатків CAD/CAE систем.

На основі архітектурного шаблону MVO побудовано 3D-графічну систему HOOPS (HOOPS/3dGS) [3], на основі якої можна проаналізувати основні складові шаблону та його особливості. HOOPS/3dGS має структуру model-view-operator, яка інкапсулює різні структури та концепції даних HOOPS/3dGS. HOOPS/MVO робить ключові можливості HOOPS/3dGS більш доступними, пропонує широкий діапазон загальних логічних рівнів застосування сприяючи ефективній розробці та швидкому створенню прототипів повнофункціональних, високопродуктивних додатків CAD/CAM/CAE систем[4].

Архітектура Model-View-Operator (MVO) відокремлює графічну інформацію від її презентації та маніпулювання. Модель являє собою дані, тоді як представлення даних відповідає представленню цієї інформації. Оператор це сукупність дій, що включають запит, створення, редагування та обробку даних. Архітектура реалізована за допомогою трьох основних класів: HBaseModel (модель), HBaseView (вид) та HBaseOperator (оператор) [2].

У MVO клас HBaseModel інкапсулює моделі, що знаходяться в Include Library. При побудові сцени спочатку створюється новий екземпляр у сегменті драйверів, який відповідає екземпляру драйвера. Клас MVO HBaseView зберігає інформацію для певного примірника драйверів та дає можливість керувати такими даними, як світло, накладення малюнка, набори виділення та розташування камери. Клас HBaseOperator та його похідні класи призначені для зміни положення камери, виділення об'єктів, або видалення різних елементів після відображення 3D-сцени.

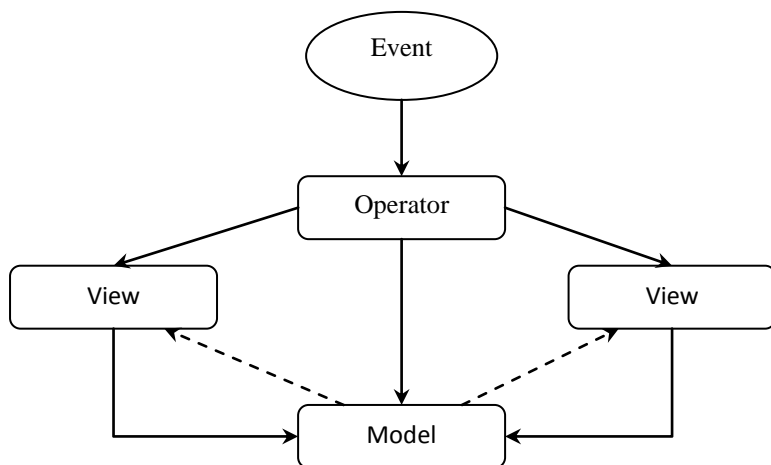


Рисунок 1 – Взаємозв'язок між окремими об'єктами в архітектурі Model-View-Operator [2]

У відповідності до шаблону MVO під час створення додатка один етап класу HDB повинен бути створений та ініціалізований під час фази ініціалізації програми. MVO підтримує кілька моделей, а також декілька виглядів. У класі HDB зберігається інформація про 3D-графіку. Щоб завантажити інформацію в базу даних, можна використати HBaseModel. Цей клас включає підтримку для завантаження різноманітних форматів файлів. Після того, як інформація знаходиться в базі даних, можна продовжувати користуватися HBaseModel, для керування 3D-об'єктами. При необхідності зберегти інформацію про об'єкт, HBaseModel включає підтримку експорту в різні формати.

Клас HBaseView керує презентацією моделей. Хоча об'єкт HBaseModel може мати кілька примірників HBaseView, кожен HBaseView асоціюється з рівномірним екземпляром HBaseModel. У базі даних HBaseView інкапсулює екземпляр сегмента драйвера. Екземпляр драйвера визначає з'єднання з OpenGL або DX9. При ініціалізації HBaseView створюється структура сегменту за умовчанням в сегменті екземпляра драйвера, який використовується для керування взаємодією з HBaseModel.

HBaseOperator є абстрактним базовим класом, який визначає інтерфейс для обробки вводу користувача та роботи на моделі або вигляді (view). Методи, визначені в цьому класі, служать основою для відображення вхідних даних користувача до логіки взаємодії. У цьому контексті об'єкт походить від HBaseOperator, а потім реалізує методи для обробки подій, таких наприклад, як рух миші або натискання клавіш.

MVO не тільки представляє основу для обробки подій користувача в класі HBaseOperator, але також надає набір попередньо сформованих класів, отриманих від HBaseOperator, які обробляють широкий діапазон взаємодій.

Імпорт та експорт даних підтримуються в MVO через архітектуру вводу-виводу файлів. В основі цієї архітектури знаходиться клас HIOManager, який обробляє всі модулі введення та виведення файлів. При ініціалізації класу HDB створюється екземпляр HIOManager.

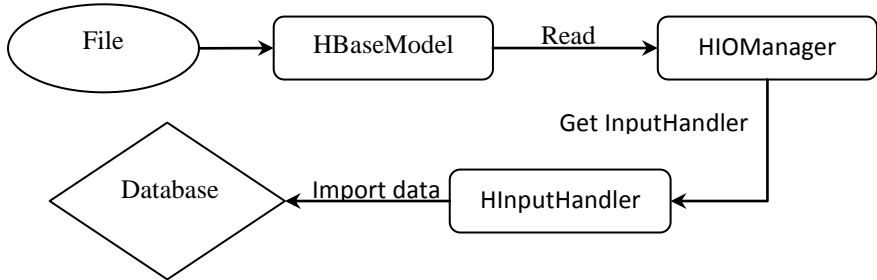


Рисунок 2 – Завантаження файла в базу даних [2]

При імпорті файлів HBaseModel:Read ініціює через HIOManager процес пошуку відповідного обробника даних для файлу. Пошук полягає, насамперед, у порівнянні з розширеннями файлів. Коли буде знайдено сумісний обробник вхідних даних, інформація буде записана в базі даних у вказаній області дерева сегментів. Експорт даних аналогічний імпорту.

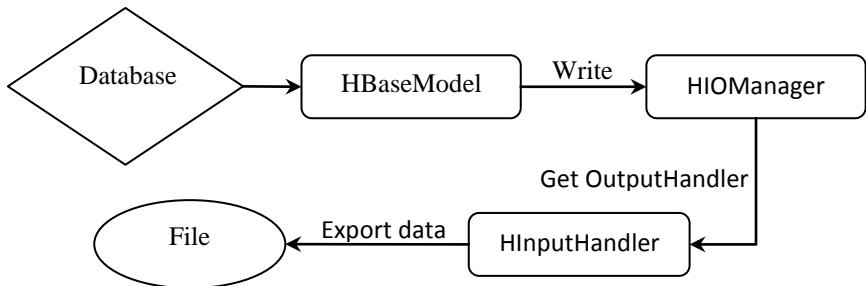


Рисунок 3 – Експорт інформації з бази даних у файл [2]

Архітектура вводу/виводу файлів HOOPS/MVO підтримує велику кількість форматів файлів.

Аналіз шаблону проектування MVO показує, що його архітектура найбільш точно підходить для використання при створенні CAD/CAE додатків, реалізує в собі методи для обробки введеної користувачем інформації, будує відповідні моделі та їх відображення, а також дозволяє не

змінюючи архітектуру системи масштабувати її, вносити зміни та доповнення до елементів системи.

Література

1. MVC для веб: проше некуда [Електронний ресурс]. – Режим доступу: <https://habr.com/post/181772/>
2. HOOPS/MVO Technical Overview [Електронний ресурс]. – Режим доступу: https://docs.techsoft3d.com/visualize/3df/latest/build/tech_overview/mvo_technical_overview.html
3. Hoops Visualize Technical Documentation [Електронний ресурс]. – Режим доступу: <http://techsoft3d.jp/developers/technical-documentation/hoops-visualize/>
4. 10 Ways to Unlock the Power of HOOPS Visualize [Електронний ресурс]. – Режим доступу: <https://www.prototechsolutions.com/hoops-visualize-application-development/>

Автоматизована система для закладу з надання косметологічних послуг

Болкун А.О., Міхалевський В.Ц.

Хмельницький національний університет

У світі коли «Інтернет» є невід’ємною частиною життя, майже, кожної людини, web-сайт є необхідною частиною для запоруки успішного бізнесу. Сайт - цілодобове представництво підприємства в Інтернеті. Від того як виглядає сайт і яку корисну інформацію знайде відвідувач на його сторінках залежить, чи стане він потенційним замовником продукції або послуг, що надаються салоном. Для досліджуваної теми «Автоматизована система закладу з надання косметологічних послуг» основна вимога залишається незмінною - її актуальність. Результати роботи не повинні припадати пилем у просторах «Інтернет», а повинні бути використані. На сайті відвідувач може ознайомитися з усіма послугами, які надає салон, забронювати місце у зручний для себе час, ознайомитися з персоналом і їхньою освітою, а також залишити відгуки про майстрів чи порадитися з ними.

Метою і призначенням цього проекту є реалізація середовища, що буде здійснювати комунікацію для взаємодії менеджерів та відвідувачів салону. А також реалізація інтерфейсу, що забезпечить можливість ознайомитися з усіма послугами, що надає салон, цінами, працівниками та бронювання місця у салоні.

Сформовані необхідні для проекрованої системи функціональні вимоги. З огляду на попередні розділи система буде використовуватися декількома групами користувачів. Приведемо список цих груп:

– організатори – група користувачів, яка створює контент сайту, кожен організатор додає нові статті та відслідковує людей, які забронювали місце у салоні;

– відвідувачі – це група користувачів, яка може бронювати місця на будь які процедури та спілкуватися з персоналом;

– гості – припускається, що це група неавторизованих відвідувачів майбутньої системи. Вони можуть здійснювати пошук процедури, але не бронювати місця на них.

Для кожної із груп користувачів існують свої функціональні вимоги. Приведемо список для групи організаторів:

а) система повинна забезпечувати можливість самостійно реєструватися новим адміністраторам та персоналу;

б) організатори можуть додавати нові статті, процедури, новини, організатори можуть відмінити новини і статті; в) при новому бронюванні місця на процедуру відвідувачем система повинна також показувати ці дані адміністратору;

г) керівник може змінювати дані;

д) проєктований програмний продукт повинен надавати доступ до системи управління тільки після авторизації за логіном і паролем.

Для користувачів майбутніх послуг необхідно врахувати наступні функціональні вимоги:

а) система повинна надавати можливість самостійної реєстрації;

б) користувач може бронювати місце у будь якій послугі, якщо є вільний час; в) користувач може відмінити бронювання;

г) користувач може ознайомитися з фотогалереєю;

д) користувач може здійснювати пошук послуг по назві або будь якому іншому критерію;

е) проєктований програмний продукт повинен надавати доступ до даних системи відвідувачам без авторизації.

Гості, або неавторизовані користувачі мають роль неавторизованого відвідувача. Дана роль надає відвідувачам сервісу можливість переглядати всі послуги, що надає салон, фотогалерею та інформацію про співробітників. Важливою для системи є також роль адміністратора. Вона висуває наступні вимоги:

а) система повинна надавати можливість додавання майбутніх адміністраторів сервісу для уже існуючих у системі користувачів, які мають роль адміністратора;

б) адміністратор може переглядати дії інших авторизованих користувачів майбутньої системи;

в) адміністратор може блокувати дії інших користувачів на форумі;

г) проєктований програмний продукт повинен надавати доступ до даних системи адміністраторам тільки після авторизації за логіном і паролем.

Щодо нефункціональних вимог до системи, то вони є уніфіковані:

- система на даному етапі повинна мати веб інтерфейс;
- система повинна бути готова до розробки мобільного клієнта;
- система повинна бути готова до розміщення на веб-хостинг хмарного типу зберігання.

Говорячи про системний підхід, можна говорити про деякий спосіб організації наших дій, такий, який охоплює будь-який вид діяльності, виявляючи закономірності і взаємозв'язки з метою їх більш ефективного використання. При цьому системний підхід є не стільки методом вирішення завдань, скільки методом постановки задач. Вважають, що "правильно поставлене питання - це вже є половина відповіді". Це якісно більш високий, ніж просто предметний, спосіб пізнання. Розроблювальний програмний продукт повинен відповідати основним системним параметрам (простота, надійність), принципам щодо інформації та основним принципам системного підходу:

- цілісність;
- ієрархічність;
- структуризація;
- системність;
- множинність.

Проектування і розробка сайту включає: затвердження первинного технічного завдання розробки сайту, визначення структурної схеми сайту - розташування розділів, контенту і навігації, веб-дизайн - створення графічних елементів макету сайту, стилів і елементів навігації, розробка програмного коду, модулів, бази даних і інших елементів сайту необхідних в проєкті, тестування і розміщення сайту в мережі Інтернет.

Згідно описаних потоків даних основною задачею даної системи є організації зв'язків салону та відвідувачів сайту та надання їм необхідного функціоналу. Зважаючи на те, що користувач програмної системи може працювати з нею із будь-якого пристрою, потрібно створити додаток, який є кросплатформним. Тому, структуру програми буде формувати архітектура «Клієнт-Сервер», де клієнт являє собою простий інтерфейс для роботи із сервером. На стороні сервера виконується вся обчислювальна логіка. Сервер буде взаємодіяти із спроектовано раніше базою даних. Отже, декомпозиція задач програми буде здійснюватися для створення серверу. Згідно планів автоматизації, вхідними даними у головному блоці є вхідна інформація про послугу від адміністратора та бронювання місця від відвідувача, а вихідними результатами є вже надана салоном послуга. Згідно моделі декомпозиції, основні варіанти роботи із програмою розділяються по ролях користувачів. Детально взаємозв'язок функціональної моделі для кожної групи користувачів, описаний у діаграмі варіантів використання, можна показати за допомогою діаграм взаємодії. За допомогою даних діаграм можна визначити

класи, які потрібно створити, зв'язки між ними, а також операції та відповідальність кожного класу.

Згідно описаної стратегії розробки, розроблювальна програмна система має складатися із двох основних модулів:

- Сервер – написаний на платформі Spring Boot; виконує всю обчислювальну логіку системи;
- Клієнт – веб інтерфейс для роботи користувача. Варто вказати, що на даному етапі ведеться розробка тільки веб-клієнта, у майбутньому будуть додатися інші клієнти веб-сервісу – мобільні або настільні додатки. Приклад взаємодії клієнта і сервера для реєстрації користувача можна побачити на рисунку 1.

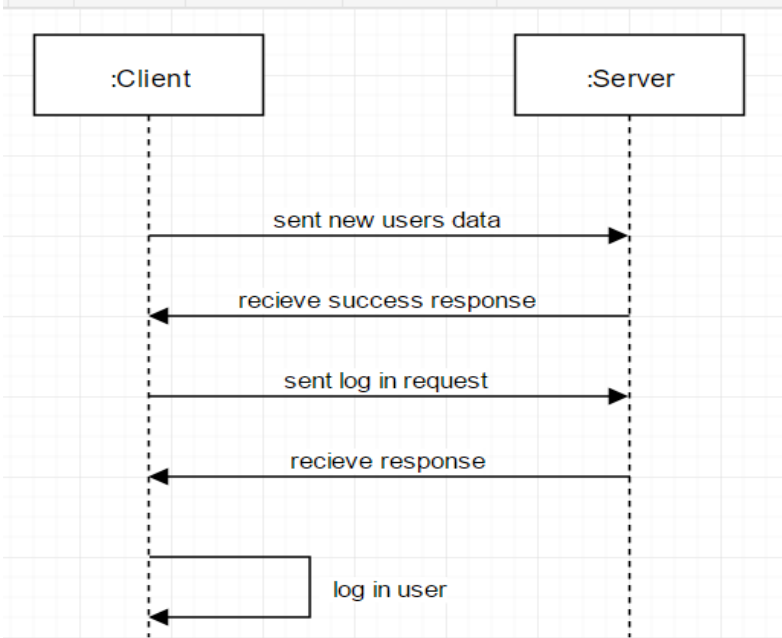


Рисунок 1 – Схема взаємодії клієнта та сервера

Мережі на базі серверів мають кращі характеристики і підвищену надійність. Сервер володіє головними ресурсами мережі, до яких звертаються інші робочі станції.

У сучасній клієнт – серверній архітектурі виділяється чотири групи об'єктів: клієнти, сервери, дані і мережеві служби. Клієнти розташовуються в системах на робочих місцях користувачів. Дані в основному зберігаються в серверах. Мережеві служби є спільно використовуваними серверами і даними. Крім того служби керують процедурами обробки даних.

Мережі клієнт – серверної архітектури мають наступні переваги:

- дозволяють організувати мережі з великою кількістю робочих станцій;
- забезпечують централізоване управління обліковими записами користувачів, безпекою та доступом, що спрощує мережне адміністрування;
- ефективний доступ до мережевих ресурсів;
- користувачеві потрібен один пароль для входу в мережу і для отримання доступу до всіх ресурсів, на які поширюються права користувача.

Отже, в даній роботі створено програмний веб-сервіс «Автоматизована система закладу з надання косметологічних послуг». В результаті розробки, на основі ряду досліджень та публікацій було проаналізовано питання впровадження та розвитку сервісу надання косметологічних послуг. Розглянуто відомі аналоги систем автоматизованих агрегаторів контенту. Основні задачі, які виконує розроблений сервіс, - це реалізація середовища, що дає змогу взаємодіяти користувачам та відвідувачам салону краси, можливість модерації сайту командою адміністраторів, а також реалізація інтерфейсу, що забезпечить процес створення визначень та оцінювання інших користувачів. Розробка даного продукту є обґрунтованою та економічно вигідною. У майбутньому сервіс можна розширити, додавши до нього низку клієнтів для різних мобільних та настільних платформ та впровадити платні послуги.

Література

1. Алистер, Коберн Быстрая разработка программного обеспечения / Коберн Алистер. - М.: ЛОРИ, 2014. - **862** с.
2. Татарчук М. І. Корпоративні інформаційні системи: Навч. посібник. – К., 2005. – 245 С.
3. https://developer.mozilla.org/ru/docs/Learn/Server-side/First_steps/Client-Server_overview

Інформаційна технологія визначення ознак обличчя, що впливають на прояви емоцій

Гащук Т.О., Бармак О.В.

Хмельницький національний університет

Із розвитком комп'ютерних технологій останнім часом науковців усе більше цікавить можливість машинного розпізнавання емоцій людей для різноманітних цілей. Наприклад таких як автоматизовані охоронні системи, що допоможуть уникнути більшості нещасних випадків у людних місцях. Виходячи із актуальності даної проблематики з'являється все більше програмних продуктів, що реалізують розпізнавання емоцій для різних цілей. Також виходить усе більше праць, що розглядають цю проблематику. Майже

усі праці спрямовані на виділення певних ознак, що допомагають визначити емоцію людини а також на розробку спеціального програмного продукту, що в динаміці або статистиці буде визначати емоцію, що в даний момент проявляється на обличчі актора. Більшість із них знаходять точки на обличчі за допомогою примітивів Хаара та виділяють такі зони обличчя як брови і рот за допомогою переводу зображення обличчя людини спочатку до полу тонового потім до бінарного і в кінці виділяють ключові точки [1-2]. Також можна виділити роботу, де точки обличчя позначені спеціальними маркерами на обличчі актора [3]. У цій роботі досліджується відео-поток та відслідковується зміщення кожного маркера. А самі емоції детектуються завдяки комбінації класичного та пірамідального алгоритмів Лукаса-Канаде. Але майже немає робіт, які вказують, як саме вони прийшли до якісних критеріїв оцінки зміщення точок і груп точок.

Також у вільному доступі можна знайти програми для розпізнавання емоцій, наприклад такі як EmoDetect [4], що визначає шість базових емоцій та визначає їх по вибірці зображень. Приклад роботи даного програмного продукту зображений на рисунку 1.

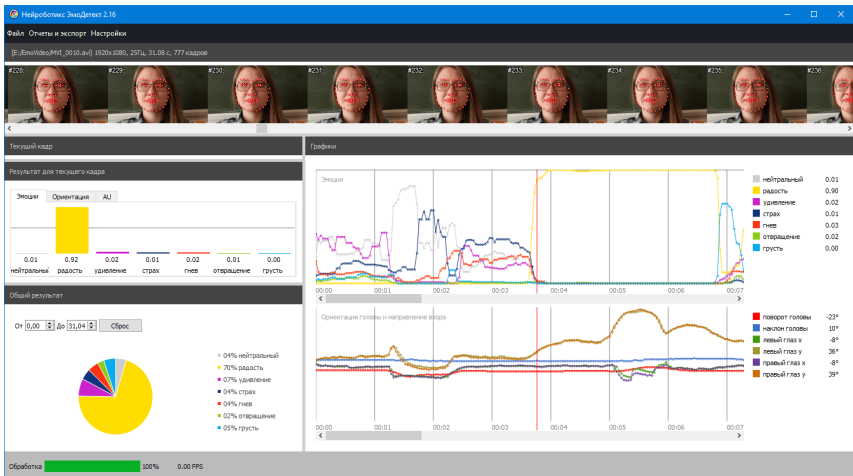


Рисунок 1 – Приклад роботи програми EmoDetect

Одним з кращих ПП розроблених на сьогоднішній час є розробка MS, що являється частиною проекту Microsoft Project Oxford.

Алгоритм програми здатний аналізувати вираз обличчя на фотографії, видаючи результат у вигляді процентного співвідношення ймовірних емоцій. Зокрема, нове ПЗ може розпізнати вираз щастя і здивування, а також емоції страху, гніву, відрази і презирства.

Окремою емоцією для програми є нейтральний вираз обличчя.

Результати мають вигляд числового значення від 0 до 1. Демонстрація подібної програми знаходиться у відкритому доступі на сайті Project Oxford [5]. На рисунку 2 зображено інтерфейс програми. Недоліком даної програми є відсутність мультиплатформенності.

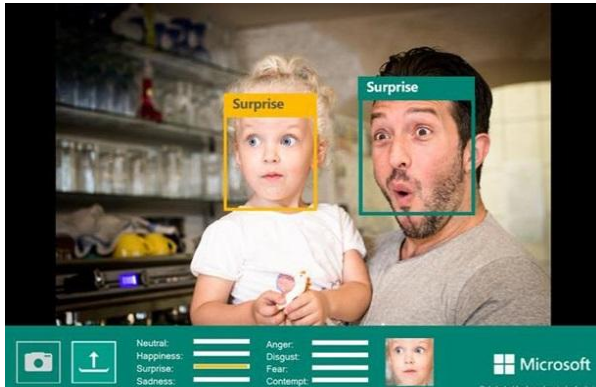


Рисунок 2 – Інтерфейс програми проекту Microsoft Project Oxford

Метою роботи є розробка інформаційної технології автоматизованого визначення емоційного стану людини, виділення основних критеріїв зміщення точок обличчя людини при емоційних проявах та мінімізація масиву значень виділених критеріїв для спрощення алгоритму детектування емоційного стану людини.

Було продовжено розробку програмного продукту, над яким працювали раніше, тому його було доповнено новими функціями. Для проведення дослідження мімічних проявів емоційного стану людини нам знадобилася форма, яка буде обраховувати кількісні зміни окремих точок, причому відкидаючи повороти та зміщення голови актора. Також нам потрібна була можливість для повернення до вибору актора для проведення іншого чи повторного дослідження. На рисунку 3 зображено схему роботи нової форми програми для обробки та дослідження зображення.

Було створено форму за допомогою засобів розробки Windows Forms, на яких будуть присутні елементи для вибору типу дослідження та актора, за даними якого буде проводитись дослідження та область виведення даних, у якій зазначено переміщення кожної із досліджуваних точок.

При написанні коду вікна програми із усіх точок, що пропонує камера Intel Real Sense [6], ми обрали лише точки на самому обличчі і лише деякі точки контуру обличчя, що дозволило пришвидшити час обрахунків програмою.

На рисунку 4 зображено приклад роботи нової форми програми

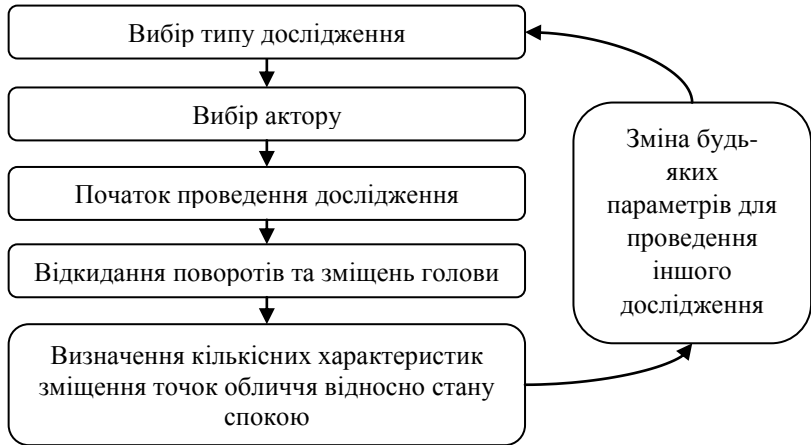


Рисунок 3 – Схема інформаційної технології для аналізу зйомки та визначення кількісних характеристик точок обличчя

Concluding Research

Type of research: Point Activity

Actor: Ольга Вальчук Start

№	Nr of point	Point name	Гнів	Горе	Задоволення	Зневага	Інтерес	Надія	Радість	Страх
1	1	Центр правого ока	21,64	0	0	9,41	13,76	71,46	3,75	10,1
2	2	Центр лівого ока	28,84	19,16	27,47	27,43	21,25	92,03	25,82	7,14
3	3	Верх правої повіки	30,92	28,73	13,74	35,82	24,62	106,47	24,69	15,97
4	4	Низ правої повіки	45,55	39,49	19,43	38,79	28,1	128,33	42,46	22,59
5	5	Права сторона правої повіки	35,47	48,84	73,97	28,22	19	28,26	39,75	15,97
6	6	Ліва сторона правої повіки	40,76	51,58	86,87	29,75	30,44	44,93	36,88	15,97
7	7	Верх лівої повіки	70,7	39,49	104,61	29,75	32,85	59,44	48,53	31,94
8	8	Низ лівої повіки	81,76	39,49	110,74	37,63	30,46	75,32	58,16	38,47
9	9	Права сторона лівої повіки	83,44	42,84	100	44,62	20,99	95,16	68,99	43,45
10	10	Ліва сторона лівої повіки	15,48	42,84	30,71	14,87	28,84	46,65	24,96	10,1
11	11	Центр правої брови	8,57	28,73	49,53	21,04	28,84	89,1	22,52	7,14
12	12	Права сторона правої брови	19,34	42,84	68,68	23,96	30,44	90,42	20,45	14,29
13	13	Ліва сторона правої брови	26,17	39,49	58,28	29,75	30,44	103,2	22,9	15,97
14	14	Центр лівої брови	27,88	42,84	43,44	29,75	25,7	117,67	33,4	10,1
15	15	Ліва сторона лівої брови	16,27	42,84	30,71	29,75	25,7	99,63	29,72	10,1
16	16	Права сторона лівої брови	16,37	34,53	27,47	21,04	25,7	60,74	29,56	10,1
17	17	Кінець носа	17,64	47,89	43,44	19,96	25	49,98	29,28	7,14
18	18	Верх носа	40,25	34,53	70,04	29,75	30,96	55,04	53,22	14,29
19	19	Низ носа	45,55	30,29	80,09	21,04	23,02	71,82	53,64	25,75
20	20	Права сторона носа	59	30,29	87,95	21,04	30,44	76,12	55,96	30,3
21	21	Ліва сторона носа	57,68	39,49	99,05	33,26	23,02	80,26	59,37	30,3
22	22	Права сторона губ	63,24	34,53	100	38,79	23,02	78,6	61,86	29,45
23	23	Ліва сторона губ	54,24	40,64	86,87	42,07	23,02	63,78	58,32	21,43
24	24	Центр верхньої губи	50	40,64	83,55	42,07	28,84	57,84	53,86	22,59
25	25	Права сторона верхньої губи	47,88	34,53	70,04	35,82	28,84	55,32	53,86	21,43

Рисунок 4 – Приклад роботи додаткового вікна програми для проведення дослідження

При проведенні досліджень для початку було з'ясовано, що тип обличчя і різні форми губ, брів та очей не впливають на результати, а ключові точки є порівняльними. За основу було взято точку кінчика носа та точки середин очей. Далі на рисунку 5 наведені відповідні фотографії із нанесеними камерою точками, що демонструють порівняльну картину точок обличчя при різній формі та розмірі головних зон обличчя.



Рисунок 5 – Відображення точок обличчя для різних форм ключових зон обличчя та «Трикутник» основних точок обличчя

Головною нерухомою точкою при відкиданні загальних зміщень голови виявиться кінчик носу, саме на цю точку ми орієнтуємося при проведенні нашого дослідження та при обрахунках.

Отже, зміщення точок для різних типів обличчя є порівняльними, а також є можливим використання отриманих даних для розпізнавання емоцій на будь-якому обличчі, що дозволяє нам при відносно невеликій вибірці застосовувати наші дослідження та їх результати для детектування емоцій на будь-якому людському обличчі або застосовувати результати досліджень для детектування емоцій засобами звичайних веб-камер із невеликою роздільною здатністю або на великій відстані.

Наступним кроком було проведено збір даних та доповнення бази даних, що вже була створена раніше, у вигляді набору фотографій акторів при прояві емоцій та виявлено ті точки, що найбільш виражено зміщуються при прояві тої чи іншої емоції.

Виходячи із даних, що отримані в ході дослідження було виділено 3 основні зони обличчя та сформовані якісні критерії зміщення точок при прояві певної емоції. Якісні критерії для певних зон обличчя наведені у Таблиці 1. Виділення даних критеріїв дозволить у майбутньому спростити алгоритм розпізнавання емоцій через зменшення кількості змінних параметрів та можливих станів, які ці параметри можуть приймати.

Таблиця 1 – Якісні характеристики зон обличчя

Емоція	Брови	Губи	Очі
Гнів	опущені	стиснутий	без змін
Горе	без змін	опущені кутики	опущені верхні повіки
Задоволення	без змін	без змін	злегка прикриті верхні повіки
Зневага	без змін	піднятий один кутик губ	без змін
Інтерес	злегка підняті	без змін	злегка привідкриті верхні повіки
Надія	злегка підняті	без змін	злегка привідкриті верхні повіки
Страх	сильно підняті	витагнуті кутики губ	підняті верхні повіки
Радість	без змін	підняті кутики губ	підняті зовнішні кутики очей

Як підсумок, було отримано масив можливих значень зон та під зон обличчя, що у подальшому допоможе спростити алгоритм детектування емоційного тану людини.

Література

1. Р.И. Ахметшин Распознавание эмоций человека на изображениях / Р.И. Ахметшин, А.П. Кирпичников, М.П. Шлеймович // Вестник технологического университета. – 2015. – Т.18, №11. – С. 160-163.
2. Ян Си Автоматическое распознавание эмоций пользователя для организации интеллектуального интерфейса / Ян Си, Алфимцев А.Н // Молодежный научно-технический вестник – сентябрь 2013. – № 09.
3. Кривонос Ю. Г. Информационная технология анализа мимических проявлений эмоциональных состояний человека / Ю. Г. Кривонос, Ю. В. Крак, А. В. Бармак, А. С. Тернов, В. А. Кузнецов // Кибернетика и системный анализ. – 2015. – Т. 51, № 1. – С. 30-39. – Режим доступа: http://nbuv.gov.ua/UJRN/KSA_2015_51_1_6.
4. Система распознавания эмоций по лицу EmoDetect. [Электронный ресурс] Режим доступа: <https://neurobotics.ru/psychophysiology/emodetect>
5. Microsoft научит компьютеры распознавать эмоции людей [Электронный ресурс]. Режим доступа: <http://korrespondent.net/lifestyle/gadgets/3588876-Microsoft-nauchyt-komputery-raspoznavat-emotsyy-luidei>
6. Технология Intel® RealSense™ Представляйте. Запечатлейте. Творите. Взаимодействуйте. [Электронный ресурс]. Режим доступа: <http://www.intel.ru/content/www/ru/ru/architecture-and-technology/realsense-overview.html>

Система підтримки та аналізу інформаційних потоків суспільної організації

Гикавчук А.В., Свірневський М.С.
Хмельницький національний університет

З розвитком інформаційних технологій використання Інтернету стало звичайною справою для більшості людей. Велика кількість галузей людської діяльності дублюються в Інтернеті, а деякі з них повністю переходять в мережу. Це виникає через те, що ресурси в реальному світі, які необхідні для ведення бізнесу, часто значно дорожчі за віртуальні. Так, наприклад, велика кількість торгових будинків комп'ютерної техніки в Китаї повністю закриваються та «переселяються» до Інтернету.

На сьогодні актуальне програмне забезпечення – це веб-програми, тому що це дуже комфортно. Не потрібно встановлювати стороннє ПЗ або наприклад щоб запустити новішу програму написану в середовищі Visual Studio потрібно оновлювати .NET Framework. А для людей, які користуються іншими операційними системами потрібно розробляти ще під кожен ОС, а на це все потрібно час. Використовуючи веб технології, можна розробити будь-яку програму яка буде працювати на всіх операційних системах та навіть телефонах. Є певні мінуси адже не всі браузерери відображають деякі стилі по різному, тому розробники повинні налаштовувати свої розробки під популярні браузерери.

В даній роботі об'єктом дослідження є діяльність благодійного фонду. Благодійний фонд – це недержавна організація, головною метою діяльності якої є здійснення благодійної діяльності в інтересах суспільства або окремих категорій осіб. Фонди об'єднують невеликі посильні пожертви в великі суми.

Метою роботи є аналіз та розробка системи, що дозволить спростити процес взаємодії користувачів з веб-сайтом(перегляд новин, перегляд звітності, здійснення пожертв) благодійної організації.

Для задоволення потреб благодійної організації на сайті необхідно реалізувати наступний функціонал:

- створення кампаній для збору коштів;
- додавання новин на сайт;
- завантаження звітів організації;
- редагування інформації про пожертви;
- створення нових адміністраторів системи;
- можливість здійснення пожертвувань користувачами веб-сайту.

Щоб реалізувати вказані вище вимоги, структурна модель такої системи повинна складатися з наступних модулів (рис. 1):

- програмний модуль роботи з адміністративною частиною;
- програмний модуль клієнтської логіки;
- програмний модуль доступу до бази даних (БД).

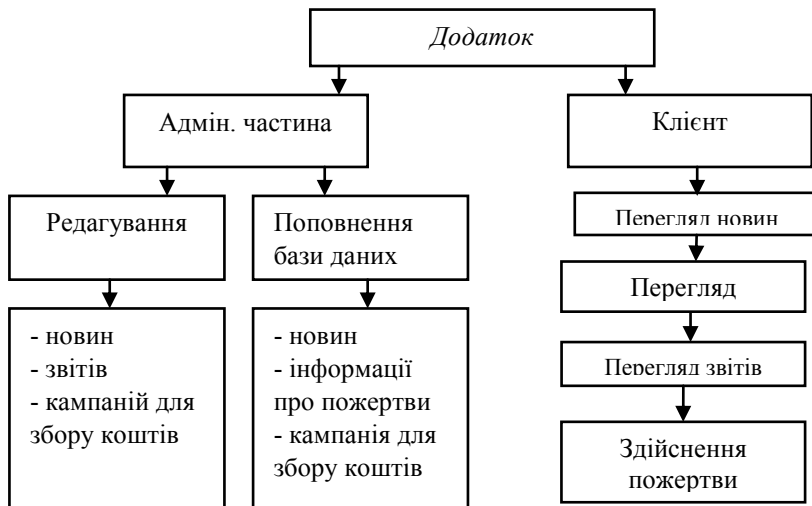


Рисунок 1 – Структурна модель веб-сайту.

Програмний модуль для роботи з інтерфейсом забезпечує коректне відображення інформації на сторінках та самих сторінок [1].

Програмний модуль головної логіки необхідний для коректної роботи веб-сторінок, здійснювати переходи по сторінкам для перегляду інформації та здійснення пожертви.

Сторінки кампаній є основою додатку. Складовою цих сторінок, зазвичай є фото, назва, опис кампанії та кнопка «Допомогти». На сторінці кампанії повинні бути деталізовані фото, необхідна сума, текстовий опис, кількість уже зібраних коштів.


«Модуль оплати» – важлива складова сайту. Він повинен бути легким у використанні. При бажанні здійснити пожертву користувач перенаправляється на сторінку здійснення транзакції(рис.2).

У додатку повинні бути й інші сторінки – сторінка новин, сторінка з формою для відправлення заявки щоб стати волонтером і безпосередньо учасником фонду.


Отже, враховуючи аналіз переваг і можливостей технологій створення сайтів, сучасну технологію мови програмування PHP та системи управління базами даних MySQL, було розроблено систему підтримки та аналізу інформаційних потоків суспільної організації, функціональна структура якого складається з веб-сайту та бази даних, дані елементи взаємодіють між собою. На основі була вибрана оптимальна архітектура клієнт-серверної системи забезпечення взаємодії і був розроблений додаток за допомогою мови програмування PHP, були використані мови HTML та CSS для правильної розмітки та опису зовнішнього вигляду сторінок.

Оплата: 200,00 Р Безопасное
соединение

Магазин: serdcagrossii
Номер заказа: 1528422872 добровольное пожертвование

Введите номер карты
5168 9337 1333 3333 

Месяц: 11 / Год: 23

CVV/CVC 

Получить квитанцию на эл. почту Оплатить

Введите адрес электронной почты вы соглашаетесь с условиями передачи информации

Рисунок 2 – Зовнішній вигляд форми пожертвування

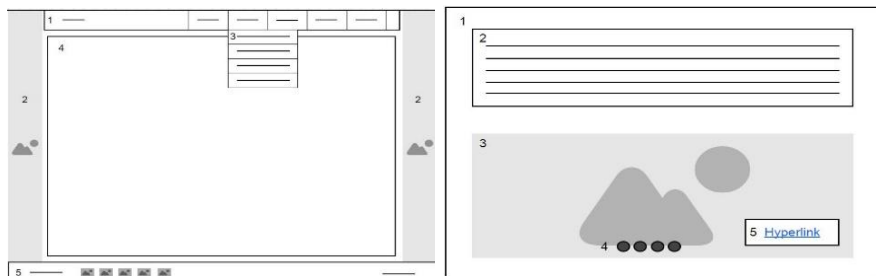


Рисунок 3 –Макети інтерфейсу веб-додатка.

Література

1. Никсон Р. Создаем динамические веб-сайты с помощью PHP, MySQL, JavaScript и CSS. 2-е изд. – СПб.: Питер, 2013.
2. Fowler M.UML distilled / M. Fowler – Boston: Addison, 1997- 2018 р.
3. Купер А. Интерфейс. Основы проектирования взаємозв'язку. 2018 р.

Інформаційна система миттєвого обміну повідомленнями в локальній мережі

Дем'янюк М.В., Багрій Р.О.

Хмельницький національний університет

В умовах сучасного інформаційного суспільства комп'ютерні технології повністю змінили способи обміну інформацією.

Ще донедавна, основним способом обміну повідомленнями була електронна пошта, але незважаючи на те, що вона стала більш мобільною, люди надають перевагу засобам комунікації з миттєвим обміном повідомлень (Instant Messaging, IM). [1]. Система миттєвого обміну повідомленнями – одна з різновидів технології комунікації в Інтернеті поряд з електронною поштою, піринговими мережами (p2p) і відео конференціями. В основу покладено принцип миттєвого пересилання повідомлень (і другого контенту) переважно між двома учасниками. Щоб використовувати подібні сервіси, необхідні лише клієнт-сервер і відповідна клієнтська програма (ІМ-клієнт).

Системи миттєвого обміну повідомленнями дають можливість передавати повідомлення, документи, файли і інші дані між користувачами сервісу, і в будь-який час отримати доступ до даних з будь-якого пристрою. Найбільшим недоліком систем миттєвого обміну повідомленнями є конфіденційність переданої інформації. Щоб вирішити питання захищеності даних користувачів використовуються різні методи шифрування даних, локальну мережу без можливості доступу до Інтернет, а також VPN-сервіси.

Метою досліджень є розробка інформаційної системи на платформі Android для миттєвого обміну повідомленнями в локальній мережі без використання доступу до Інтернет. Локальна мережа забезпечує високий рівень захисту конфіденційної інформації і захист від несанкціонованого доступу в роботу системи.

На основі даної інформаційної технології необхідно створити програмний додаток для забезпечення швидкого обміну повідомленнями в локальній мережі. Так, як більшість користувачів активно використовує смартфони, планшети, смарт годинники то, розробка додатку буде проводитися на платформі Android.

В рамках проведених досліджень було проаналізовано системи миттєвого обміну повідомленнями на платформі Android:

1. Telegram – додаток має великий перелік функцій, таких як: відправка/отримання повідомлень, фото, документів. Споживає мінімум трафіку, безкоштовний(не містить рекламу), листування зберігається на сервері, автосинхронізація на всіх пристроях, є можливість редагувати повідомлення. Також у месенджері є групові чати, секретні чати, голосові дзвінки, стікери, боти [2].

2. Sotfros LAN – це безкоштовний месенджер для локальної мережі,

який має невеликий перелік функцій, таких як: відправка/отримання повідомлень, фото [3].

Також в рамках проведених досліджень було проаналізовано протоколи передачі даних:

1. TCP/IP – один з основних протоколів передачі даних в мережі, призначений для управління передачею даних. Основні переваги: висока швидкість обміну даними, надійність [4].

2. XMPP – протокол обміну повідомленнями та інформацією, раніше відомий як Jabber, оснований на XML. Основні переваги: розподіленість, високий рівень безпеки, використання децентралізованої архітектури (відкрити для всіх користувачів) [5].

Отже на основі аналізу аналогів програмного забезпечення та протоколів передачі даних для миттєвого обміну повідомленнями можна зробити висновок, що додаток повинен бути безкоштовним, простим і водночас зручним у користуванні, мати швидку та надійну систему доставки повідомлень.

Враховуючи зазначені вимоги розроблено інформаційну систему (ІС), що дає можливість виконувати основні функції системи миттєвого обміну повідомленнями, та забезпечує його стабільну роботу (рис.1).



Рисунок 1 – Схема роботи інформаційної системи

Інформаційна система складається з наступних модулів:

1. Модуль авторизації;
2. Модуль доставки повідомлень.

Однією з складових інформаційної системи миттєвого обміну

повідомленнями є модуль авторизації [6]. Для роботи з модулем авторизації використовується OAuth2 API. При реєстрації кожен користувач отримує індивідуальний ідентифікатор. Вся інформація про ідентифікатор, логін, пароль зберігається в базі даних на локальному сервері.

Принцип його роботи полягає в тому, що при введенні користувачем логіну і паролю, сервер перевіряє введені дані, якщо дані не вірні сервер видає помилку про невірно введені дані і перенаправляє на вікно авторизації.

Якщо дані вірні, то проводиться авторизація і перевірка на коректність введених даних. Принцип роботи модуля зображено на рисунку 2.

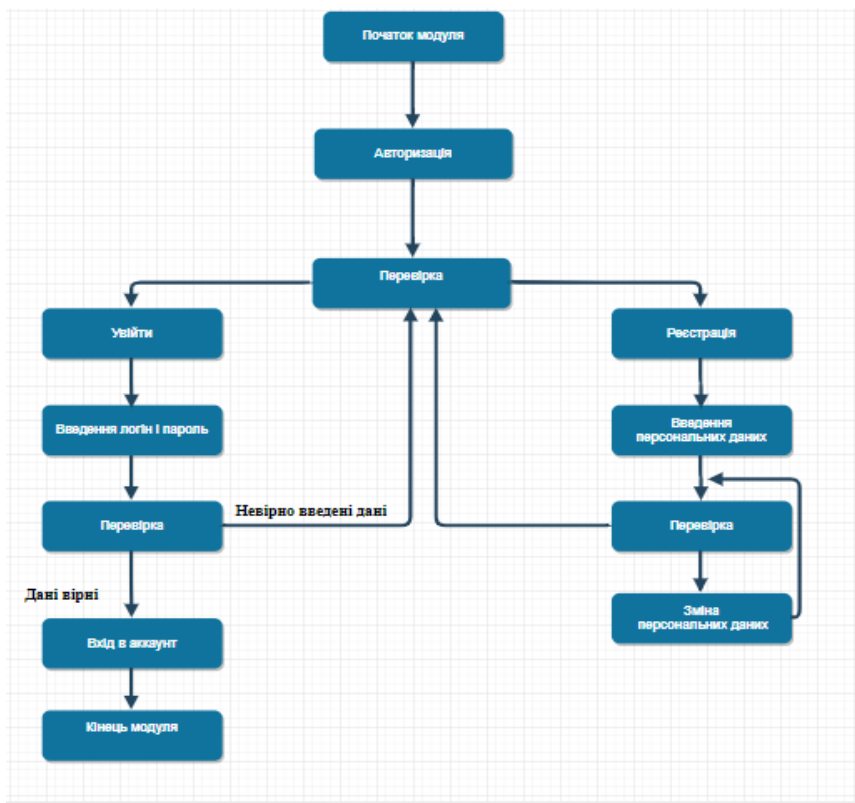


Рисунок 2 – Модуль авторизації

Для безпосередньої роботи з повідомленнями, використовується модуль доставки повідомлень. Відправка і отримання повідомлень відбувається за допомогою TCP/IP. Взаємодія між клієнтом і сервером є асинхронною і протікає шляхом передачі так званих пакетів даних.

Створюючи нове повідомлення, ми передаємо запитом на локальний сервер: ідентифікатор автора повідомлення, ідентифікатор одержувача повідомлення, текст повідомлення.

Протокол TCP розбиває інформацію на частини (пакети) і нумерує всі ці частини, щоб при отриманні можна було правильно зібрати інформацію. Після цього сервер отримує розбите на пакети повідомлення, перевіряє цілісність пакетів і потім за допомогою протоколу IP всі частини передає одержувачу, де за допомогою протоколу TCP перевіряється, чи всі частини отримані. Після отримання всіх частин TCP розміщує їх в потрібному порядку і збирає в єдине ціле.

Отже, інформаційна система для миттєвого обміну повідомленнями в локальній мережі дозволить відправляти та отримувати текстові повідомлення без використання доступу до Інтернет. Це дозволить використовувати додаток в різних корпоративних локальних мережах для підвищенні інформаційної безпеки переданих даних.

Основними перевагами розробленої інформаційної системи для миттєвого обміну повідомленнями в локальній мережі є:

- Можливість користуватися додатком, без активного з'єднання з Інтернетом;
- Лише користувачі, що знаходяться в даній локальній мережі, зможуть отримати доступ до обміну повідомленнями;
- Передані дані не залишають локальну мережу, тому система гарантовано буде вільна від спаму з зовнішньої мережі.

Інформаційну систему можна розширити шляхом розширення функціоналу месенджера, доповненням нових функціональних можливостей таких як голосові та відео дзвінки.

Література

1. Instant Messaging [Електронний ресурс]. – Режим доступу: <https://www.techopedia.com/definition/402/instant-message-im>
2. Telegram [Електронний ресурс]. – Режим доступу: <https://play.google.com/store/apps/details?id=org.telegram.messenger&hl=ru>
3. Softfros LAN [Електронний ресурс]. – Режим доступу: <https://messenger.softfros.com>
4. TCP/IP [Електронний ресурс]. – Режим доступу: <https://ua.wikipedia.org/wiki/TCP/IP>
5. XMPP [Електронний ресурс]. – Режим доступу: <https://ua.wikipedia.org/wiki/XMPP>
6. Аутентифікація через OAuth2 [Електронний ресурс]. – Режим доступу: <http://developer.alexanderklimov.ru/android/theory/oauth.php>

Пошук змісту в текстовій інформації

Джурабаєв О.В., Бармак О.В., Манзюк Е.А.
Хмельницький національний університет

Кожного дня мільйони людей з усього світу так чи інакше стикаються з текстовою інформацією. З розвитком інформаційних технологій, включаючи соціальні мережі, форуми та інші ресурси, виникає потреба у класифікації та систематизації отриманої інформації, і для часткового вирішення цієї проблеми існують готові програмні рішення. Мова – непроста річ, і сама ідея реалізації алгоритму розуміння людської мови є цікавим нетривіальним завданням, тому кожна технологія чудово підходить для вирішення конкретного, специфічного, вузького кола завдань. І хоч з 90-х років минулого століття і була реалізована велика кількість програмних продуктів для дослідження тексту, як новачками-аматорами, так і професіоналами, актуальність пошуку сенсу в текстовій інформації стоїть гостро й до сьогодні.

Розпізнавання змісту в тексті, в ідеальному випадку, потребує розробки синтаксичного аналізатора мови, а також ряду програмних лінгвістичних компонентів. Окрім цього, для коректних результатів часто потрібно мати навчальну вибірку або базу, а також проводити ресурсозатратні обрахунки. Це пов'язано з тим, що результати проведення аналізу тексту можна умовно поділити на дві групи: до першої групи віднесемо випадок для якого достатньо наближених або приблизних результатів; в іншому випадку необхідно отримати на виході точні, конкретні результати [1].

Метою роботи є розробка інформаційної технології для пошуку змісту в тексті, яка з мінімальними затратами часу та потужності ЕОМ повертає коректний результат для певного кола завдань у разі дотримання обмежень вхідних даних.

В роботі запропоновано підхід для побудови «розпізнавача» сенсу, який не потребує ні навчальної бази, ні глибокого машинного аналізу тексту і повертає наближений результат. Підхід полягає в нормалізації тексту, побудові векторів амплітуди та фази, після чого будуються графіки залежностей обрахованих параметрів та з'являється можливість аналізувати текст візуально. Також описані результати експериментів по розпізнаванню змісту в підготовлених текстах.

Вхідними даними для застосування технології є текстовий контент, який має певний сенс. Це може бути, наприклад, наукова стаття, кулінарний рецепт або якась новина чи повідомлення. Тематика інформації обмежується уявою користувача і може бути довільною. Важливим є лише те, що ця інформація повинна бути у вигляді текстового файлу.

В основі функціонування інформаційної технології пошуку сенсу в

тексті є дослідження тексту як скінченної множини слів. Схематична робота програми зображена на рисунку 1. Розглянемо кожен пункт детальніше.



Рисунок 1 – Схема роботи інформаційної технології для пошуку сенсу в текстовій інформації

Вибір текстового файлу – передбачає подання на вхід користувачем файлу, що містить необхідну для аналізу текстову інформацію.

Фільтрація включає в себе видалення з тексту розділових знаків, html-тегів та усіх інших символів, які відсутні в алфавіті. Нормалізація – приводить кожне слово до нормальної форми [2], наприклад, для слова «технологією» нормальна форма – «технологія». Після нормалізації текстовий файл є множиною слів, приведених до нормальної форми.

В свою чергу, детальний аналіз складається з кількох етапів, зображених на рисунку 2.



Рисунок 2 – Схема роботи детального аналізу текстової інформації

Для кращого розуміння суті здійснених у розрахунках величин, проведемо аналогію з таким фізичним явищем, як сигнал. Так само, як сигнал є сукупністю змін фізичної величини [3], множини слів можна представити як сукупність термів, які утворюють скінченний потік даних. Звідси слідує, що існує можливість дослідити кожен терм у тексті на предмет частоти, фази і амплітуди.

Позначимо множину слів як W , слово як w , множину унікальних термів як S , а терм як s .

Частота терму – це кількість його входжень у множину слів:

$$F(s_i) = \underset{w_i \in W}{\text{count}}(w_i), s \in S, i = 1, \dots, |S| \quad (1)$$

Фазова частота терму – кількість входжень терму в підмножину множини слів. Нехай W' – підмножина множини слів, тоді:

$$F_\omega(s_i) = \underset{w_i \in W', W' \subseteq W}{\text{count}}(w_i), s \in S, i = 1, \dots, |S| \quad (2)$$

Фаза терму – індекс, при якому його фазова частота має максимальне значення:

$$\omega(s_i) = \underset{s_i \in S}{\text{index}}(\text{argmax}(F_\omega(s_i))), i = 1, \dots, |S| \quad (3)$$

Амплітуда терму – максимальне значення зміщення або зміни кількості входжень терму від його середнього значення, іншими словами розсіювання конкретного терму в множині слів:

$$A(s_i) = \sum_{s_i \in S} \left| \underset{s_i \in S}{\text{index}}(s_i) - \omega(s_i) \right| \quad (4)$$

Підставивши $\omega(s_i)$, отримаємо:

$$A(s_i) = \sum_{s_i \in S} \left| \underset{s_i \in S}{\text{index}}(s_i) - \underset{s_i \in S}{\text{index}}(\text{argmax}(F_\omega(s_i))) \right| \quad (5)$$

Результати розрахунків по вищевказаним формулам записуються в файл, після чого можуть бути використані для візуалізації аналізу тексту.

Для прикладу візьмемо статтю «Приготування їжі як вид мистецтва». Вхідний файл містить 126 слів, з них унікальних, тобто термів – 80.

Нижче зображений точковий графік залежності амплітуди, частоти та її фази на прикладі вхідної статті. Кожна точка цього графіку є унікальним термом без повторень.

З трьохмірного представлення тексту (рис. 3) можна виділити кілька груп або кластерів. Якщо згрупувати множини точок частоти і амплітуди, утвориться графік так званої «густини» термів, приклад якого зображений на рисунку 4. Тут чітко можна розрізнити 4 кластери. Розглянемо їх.

Вектор термів, які мають велику амплітуду і малу частоту – рівномірно розподілені по всьому тексту та рідко зустрічаються. В даному випадку це слова: «для», «определённый», «необходимый», «при», «себя».

Вектор термів, які мають малу амплітуду та малу частоту – зустрічаються рідко, але концентровано. До нього входять наступні слова: «технология», «в», «магазин», «от», «сам», «кулинария», «называются», «по», «от», «или».

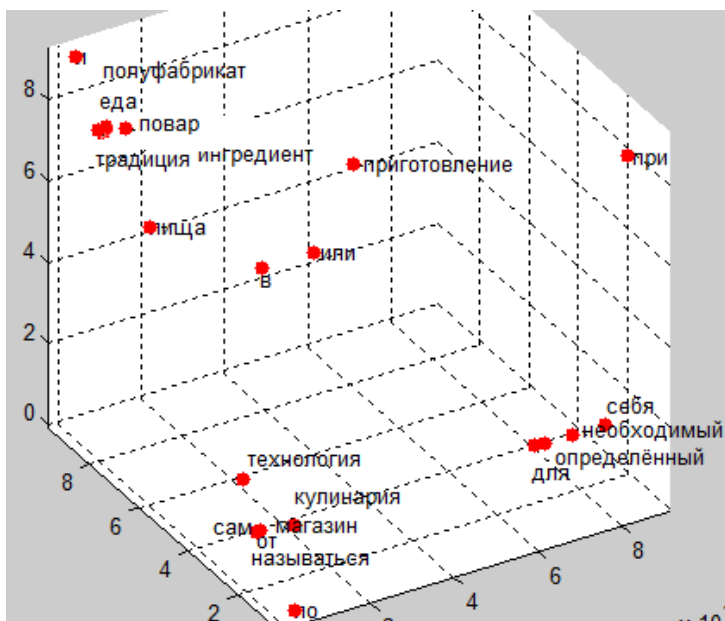


Рисунок 3 – Точковий графік обробленого вектору слів

В ході ряду експериментів було виявлено, що терм або група термів, яка є рівновіддаленою від інших груп та має помірну амплітуду і частоту, утворює вектор контексту та її можна вважати головною тематикою текстової інформації. В даному випадку це один терм – «приготовление». Хоча за певних умов ця закономірність не справджується, особливо, коли вхідна текстова інформація не має чіткої тематики.

Слова «традиция», «полуфабрикат», «ингредиент», «кухня» та всі інші, які залишилися, мають високу частоту та малу амплітуду, що говорить про те, що вони часто зустрічаються на відносно малих ділянках тексту.

Дослідження тексту не обмежується побудовою «плоских» графіків, – в деяких випадках може виявитись корисною поверхня. Наприклад, при дослідженні подібних за тематикою статей і побудові їх поверхонь можна дослідити точки перетину, – в широкому розумінні це буде ядро класифікатора або набір слів, котрі найкраще характеризують тему вхідного тексту.

Варто зазначити, що кращі результати можна отримати тільки в тому випадку, коли текстова інформація несе зміст з певної конкретної категорії. При «суміші» різноманітних тем в одній статті отримуються результати, за яких не вдається зробити коректний висновок щодо змісту тексту.

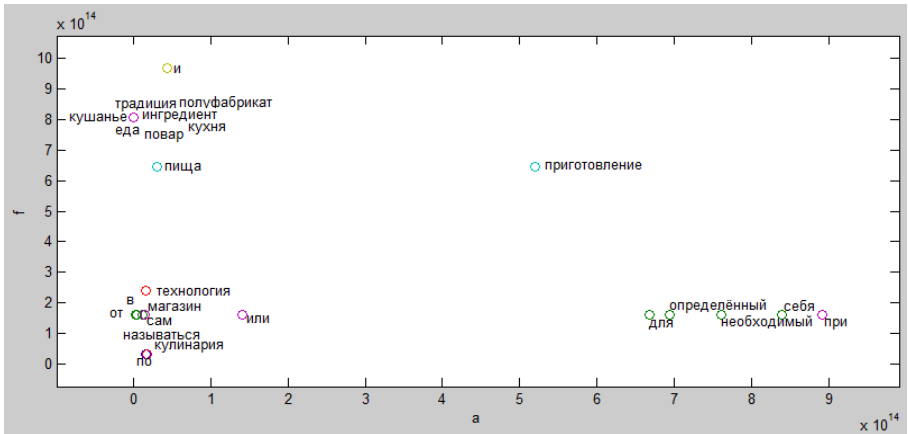


Рисунок 4 – Графік густини термів вектору слів

Таким чином, інформаційна технологія для пошуку змісту в текстовій інформації дозволяє дослідити текст графічно та забезпечує повернення коректного результату при введенні коректних вхідних даних для скінченної множини задач.

Також інформаційна технологія може бути використана як «грунт» для роздумів і подальших досліджень текстової інформації у новому, незвичному для тексту форматі.

Література

1. Єрмаков А.Є., Поляков П.Ю. «Статистична модель для розпізнавання сенсів у текстах іноземною мовою з навчанням на прикладах з паралельних текстів»: Підручник / А.Є. Єрмаков. – Москва, 2017. – 397 с.
2. Нормалізація і лематизація тексту з використанням тезаурусу [Електронний ресурс]. – Режим доступу: http://www.solarix.ru/for_developers/docs/text-normalization.shtml
3. Сигнал [Електронний ресурс]. – Режим доступу: <https://uk.wikipedia.org/wiki/Сигнал>

Інформаційна система для роботи з сервісами хмарних сховищ

Колошинський Д.В., Багрій Р.О.

Хмельницький національний університет

З розвитком світу інформаційних технологій, мережі Інтернет і зростаючим потоком інформації виникла необхідність в розвитку технологій, пов'язаних зі швидкістю обміну інформацією, забезпечення доступу до неї в режимі 24/7, а також її безпеку і захист. Одним з варіантів подібних

технологій – хмарні технології (Cloud Technology) [1]. За останнє десятиліття широке поширення в сфері хмарних технологій отримало хмарне сховище даних, що є доступним і ефективним способом зберігання інформації. Хмарне сховище даних – це сховище файлів в глобальній інформаційній мережі, в якому принцип зберігання забезпечується за рахунок численних файлових серверів, представлених для користування клієнтам. Завдяки цьому способу зберігання інформації потреба в традиційних носіях інформації, таких як «флеш-карта», «жорсткий диск», «CD або DVD-диск», поступово починає відпадати.

Хмарні сховища даних дають можливість зберігати документи, файли і інші дані в Інтернет, і в будь-який час отримати доступ до даних з будь-якого пристрою. Головним недоліком все ж є безпека інформації, що зберігається, особливо щодо конфіденційних і приватних даних. Для вирішення цих проблем можуть бути використані хмарні шлюзи [2], які використовуються для більш зручного надання хмарного ресурсу клієнту. Шлюзи хмарного зберігання даних дозволяють поліпшити захист даних, пропускну можливість, швидкість передачі великих обсягів даних. При одночасному використанні декількох різних сервісів ускладнюється навігація, що уповільнює роботу користувача з даними.

Метою дослідження є розробка інформаційної системи, що дасть можливість користувачам, маючи не одне хмарне сховище, організувати роботу з хмарними сховищами та сервісами в одному додатку.

На основі даної інформаційної системи необхідно створити програмний додаток для забезпечення файлових операцій з хмарним сховищем в єдиному файловому менеджері. З врахування того, що зараз активно використовуються мобільні пристрої, такі як планшети і смартфони, розробка такого додатку є актуальною для платформи Android.

В рамках досліджень зроблено аналіз аналогічних мобільних додатків з можливістю роботи з файлами під операційну систему Android від Google:

1. Primadesk (Web, iOS, Android) – додаток надає значний список сервісів, таких як файлові та поштові сервіси, фотохостинги та інші. Однак при безкоштовному використанні кількість підключених сервісів не може перевищувати п'яти[3].

2. ES Провідник (File Manager) – це безкоштовний провідник, який підтримує всі типи файлів (додатки, музика, фото) для локального та мережевого використання[4].

Основним недоліком розглянутих файлових менеджерів є обмежений функціонал безкоштовних версій та повна відсутність підтримки хмарних технологій або обмежена підтримка одного хмарного сховища.

Враховуючи зазначені недоліки розроблено інформаційну систему (ІС), що дає можливість виконувати основні операції з файлами, та забезпечує роботу з хмарними сховищами (рис.1).

Інформаційна система складається з наступних модулів:

1. Модуль обробки відповіді від сервера;

2. Модуль авторизації;
3. Модуль взаємодії з сервером хмарного сервісу.

Модуль обробки відповіді від сервера відповідає за отримання списку файлів хмари, в цьому випадку додаток "спілкується" з сервером за допомогою HTTP запитів [5]. Для цього запускається модуль взаємодії з сервером, який формує GET або POST запит для сервера, а у відповідь отримує json об'єкт. Принцип роботи модуля зображено на рисунку 2.



Рисунок 1 – Схема інформаційної системи

Для безпосередньої роботи з хмарними сервісами, надається відповідний для кожного сервісу Sync API [6], який дозволяє працювати з додатками з локальним сховищем на пристрої.

Файли в об'єктному сховищі супроводжуються мета-даними, які дозволяють обробляти ці файли як об'єкти: документи, відеозаписи, проекти, фотографії. Це все застосовується в самих різних додатках. Для взаємодії з хмарним об'єктним сховищем використовується програмний інтерфейс .Win. Серіалізують дані в JSON, зберігає в файли і синхронізує з хмарою. При цьому можна обмінюються даними між різними пристроями, в тому числі і між різними платформами. Також можна працювати з даними навіть і без інтернет-з'єднання, так як при підключення до мережі інтернет відбудеться автоматична синхронізація з сервісами.

Для роботи з хмарними сховищами запускається модуль авторизації, суть якого полягає в тому, що додаток спочатку шукає пов'язаний з цією хмарою акаунт і пропонує використовувати його (в разі знаходження такого), інакше перенаправляє на вікно авторизації. Далі проводиться авторизація і перевірка на коректність введених даних.

Sync API значно спрощує інтеграцію в додатки і роботу з файлами. Він уже сам по собі вирішує всі питання і проблеми, які пов'язані з хмарним

зберіганням файлів.

Отже, запропонована інформаційна система для роботи з сервісами хмарних сховищ дозволяє здійснювати обробку файлів для декількох хмарних сховищ, організувати зручну навігації по ним в межах одного мобільного додатку на смартфоні.

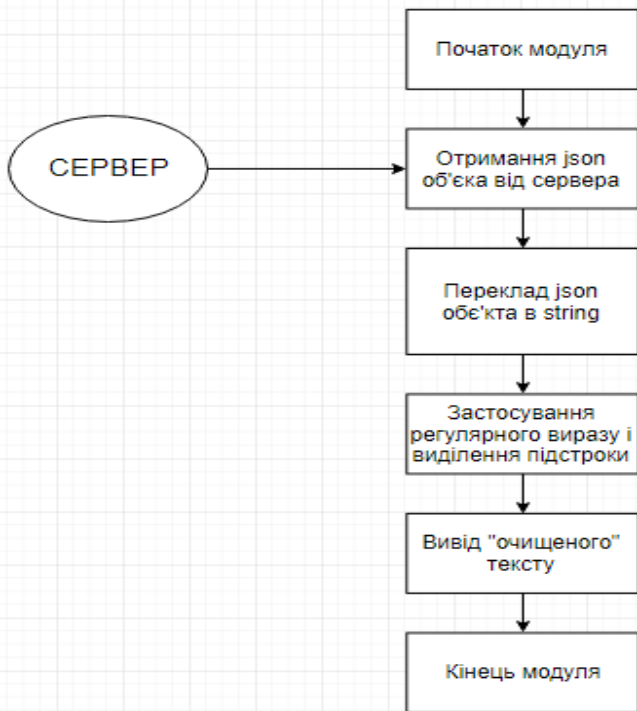


Рисунок 2 –Модуль обробки відповіді від сервера

Інформаційну систему можна розширити шляхом додавання інших хмарних сервісів або доповнення нових функціональних можливостей таких як синхронізація файлів між хмарами.

Література

1. Computation and Storage in the Cloud: Authors: Dong Yuan Yun Yang Jinjun Chen, 1st February 2013. –23 с.
2. Удаленное хранение данных и доступ к ним [Електронний ресурс]. – Режим доступу: <https://www.ibm.com/cloud-computing/ru-ru/learn-more/what-is-cloud-storage/>
3. Unifyle product Overview [Електронний ресурс]. – Режим доступу: <https://www.unifyle.co/product-overview.html>

4. ES Проводник [Електронний ресурс]. – Режим доступу: <https://play.google.com/store/apps/details?id=com.estrongs.android.pop&hl=ru>
5. GET/POST Запросы [Електронний ресурс]. – Режим доступу: [https:// developer.android.com/training/volley/simple.html](https://developer.android.com/training/volley/simple.html)
6. Documentation android[Електронний ресурс]. – Режим доступу: <https://developer.android.com/training/sync-adapters/>

Експертна система закладів громадського харчування

Кухарчук М.Н, Пасічник О.А., Скрипник Т.К
Хмельницький національний університет

Актуальність. Інформація оточує нас зі всіх сторін. У всіх установах та закладах зберігається велика кількість інформації. Для зручного її застосування використовують бази даних. Але не можна обмежитись тільки зберіганням інформації, для її обробки та застосування розробляються програмні продукти на основі систем керування базами даних. Тому перспективним є розвиток саме цієї ланки у розробці програмних продуктів.

Для роботи з даними використовують відповідні програмні продукти. Такими продуктами є СКБД [1], які почали використовуватись після появи ПК. СКБД використовуються у всіх областях, починаючи від навчальних закладів завершуючи державними установами. Однією з переваг сучасних СКБД є низькі вимоги до апаратної частини ПК. Крім того кожна нова версія СКБД розробляється з думкою про користувача. Інтерфейс стає більш інтуїтивним, а набір функцій розраховується таким чином, щоб пересічний користувач легко зміг зрозуміти суть.

Закладами громадського харчування користуються багато людей [2]. Усі вони мають різні смаки, полюбляють страви тієї чи іншої кухні, деякі інгредієнти можуть фізично не переносити. Крім того за даними Головного управління статистики у 2017 році було зареєстровано 2477,5 тисяч випадків уперше зареєстрованих захворювань. Часто при деяких видах захворювань людям категорично забороняється вживати ті чи інші продукти. Наприклад, при захворюваннях системи кровообігу (зазвичай проблеми з серцем), лікарі не радять вживати жири, холестерин, цукор. Згідно статистичних даних тільки у 2017 році зареєстровано 113,9 тисяч випадків серцево-судинних захворювань. Всі ці люди повинні обмежувати себе у вживанні не рекомендованих продуктів.

Робота закладів громадського харчування потребує використання великої кількості інформації і правильного її застосування. Інформація, яку зберігають БД закладів громадського харчування не закінчується на рецептурах. Зазвичай заклади високого рівня зберігають дані про вподобання клієнта, його смаки, хвороби тощо. Тому для автоматизації та оптимізації

роботи закладів громадського харчування доцільним є застосування інформаційних технологій.

На даний момент для автоматизації та оптимізації роботи закладів громадського харчування існує декілька програмних продуктів. Одними з рішень є програмні продукти «iОбщепит» від компанії 1С [3] та RestArt: АРМ Повара від компанії 1С-Rarus [4]. Застосунки такого типу автоматизують роботу одразу декількох відділів закладу громадського харчування. Вони дозволяють швидко оформити замовлення, проконтролювати його видачу і швидко розрахувати клієнта. Недоліками таких програмних продуктів є відсутність можливості формування замовлення з урахуванням персональних особливостей клієнта в частині, що стосується здоров'я.

Виникає проблема, що заклади громадського харчування ігнорують потреби хворих людей у класифікації їх продукції відповідно до типів захворювань. Класифікацією об'єктів у вузькій предметній області займаються експертні системи.

Експертна система – комп'ютерна система, яка може частково або повністю замінити спеціаліста-експерта в рішенні проблемної ситуації [5]. В інформатиці експертні системи розглядаються разом з базами знань, як моделями поведінки експертів в окремій області знань з використанням процедур логічного виводу та прийняття рішень, а бази знань – як сукупність фактів і правил логічного висновку в обраній предметній області.

База знань складається з правил аналізу інформації від користувачі стосовно конкретної проблеми [6]. Експертна система аналізує ситуацію і, в залежності від направленої ЕС, дає рекомендацію стосовно рішення проблеми – рис. 1.



Рисунок 1 – Стратегії отримання рішень

Як правило, база знань містить факти (статистичні дані про предметну область) та правила – набір інструкцій, використовуючи які до конкретних фактів можна отримати нові факти.

Зазвичай факти в базі знань описують ті явища, котрі є постійними для даної предметної області. Характеристики, значення котрих залежать від умов конкретної задачі, експертна система отримує від користувача в ході роботи і зберігає їх в робочій пам'яті.

Бази знань експертних систем створюються за допомогою трьох груп людей.

1. Експерти тієї проблемної області, до котрої відносяться задачі які вирішуються експертною системою.

2. Інженери по базам знань, які є спеціалістами по розробці інтелектуальних інформаційних систем.

3. Програмісти, які здійснюють реалізацію експертної системи.

Експертна система може функціонувати в двох режимах.

1. Режим вводу знань – в цьому режимі експерт за допомогою інженера по базам знань за допомогою редактора вводить відомі йому відомості про предметну область в базу знань експертної системи.

2. Режим консультації – користувач веде діалог з експертною системою, повідомляючи їй відомості про поточну проблему і отримує рекомендації від ЕС.

Експертні системи закладів громадського харчування можуть на основі даних експертів підібрати оптимальний набір страв для клієнтів зважаючи не тільки на смакові вподобання клієнта, але і на дієтологічні вказівки.

Об'єкт дослідження. Процес формування споживчого меню в закладах громадського харчування.

Предмет дослідження. Моделі, методи, підходи та засоби для побудови експертної системи закладів громадського харчування.

Результати роботи. Вперше розроблено технологію, яка дозволяє при формуванні споживацького меню у закладі громадського харчування врахувати персональні особливості клієнта в частині, що стосується стану його здоров'я.

Практичне значення одержаних результатів. Розроблено експертну систему підбору меню страв у закладах громадського харчування. В результаті було створено застосунок експертної системи. Застосування технології дає можливість формувати замовлення з урахування стану здоров'я клієнта. При формуванні меню враховуються не тільки смакові вподобання, а й дієтологічні вказівки.

Мета роботи полягає у розробці інформаційної технології підбору споживацького меню у закладах громадського харчування, яка дозволяє врахувати персональні особливості клієнта в частині, що стосується стану здоров'я. Для досягнення поставленої мети визначені наступні завдання:

1. розробити інформаційну технології для роботи з базою захворювань;
2. розробити інформаційну технологію для роботи з базою страв;
3. розробити метод створення правил, систему оцінювання та інформаційну систему для роботи з експертною базою знань;
4. виконати алгоритмічну та програмну реалізацію експертної системи для підтвердження її життєздатності.

Методи дослідження, застосовані для вирішення поставлених завдань: для реалізації інформаційної системи для роботи з базами страв та захворювань – методи формування оптимальної структури бази даних та параметричного пошуку; для формування бази знань – методи анкетування статистичної обробки даних; для реалізації експертної системи – методи багатокритеріальної оптимізації; для реалізації інформаційної технології – методи проектування інформаційних систем та об’єктно-орієнтований підхід.

Розроблюваний програмний продукт націлений на декілька груп користувачів – працівники кухні, експерти, користувачі. Робота з ними має проходити у декілька етапів, як показано на рис. 2. На кожному етапі описується стадія, починаючи від занесення довідкової інформації – до режиму підбору.

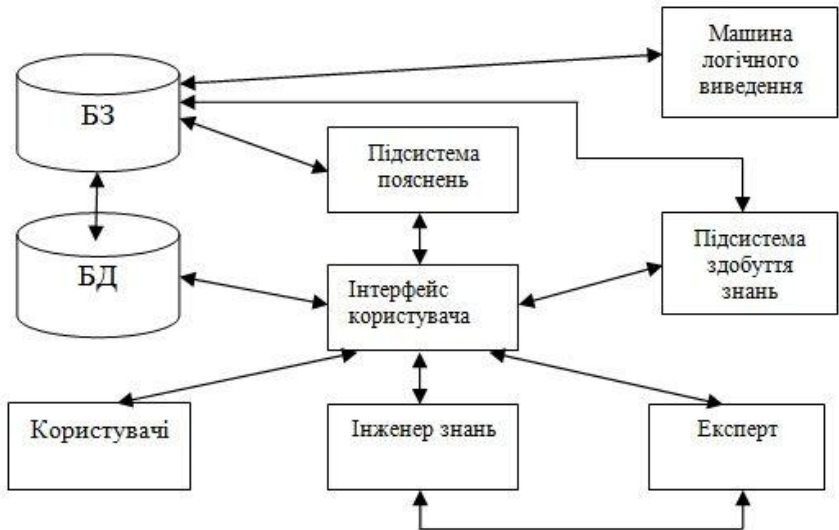


Рисунок 2 – Принцип роботи застосунку

На етапі внесення в ЕС знань експертів в залежності від правил формується список коефіцієнтів за яким ЕС буде консультувати користувачів. Рядок коефіцієнтів буде мати наступний вигляд – «Стіл, 0.01, 1,0.9,0.01, 2,0,0.01, 4,1,0.001, 7,0.9,0.03». Перший елемент – це елемент, до

якого застосовується ЕС. Після коми вказується апіорна ймовірність даного результату. Вона показує, що в разі відсутності додаткової інформації – результат події дорівнюватиме 1%. Далі через кому йде ряд повторюваних груп з трьох елементів. Перший елемент кожної групи – це номер відповідного запитання (симптому, свідoctва, ознаки). Наступний елемент – це ймовірність відповіді «так» на питання 1, якщо можливий результат вірний (P_y – це P «yes»). Третій елемент показує ймовірність того, що можливий результат невірний. Згідно прикладу, для першого запитання існує 1% ймовірності того, що якщо ми отримаємо відповідь «так» на перше запитання, результат буде не вірний.

На етапі режиму консультації/модерації програмний продукт готовий до роботи з користувачами. Принцип роботи бази знань заснований на математичних розрахунках, які виконуються за формулою повної ймовірності [7]:

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum P(B_i)P(A | B_i)} \quad (1)$$

Формула повної ймовірності дозволяє переоцінити ймовірність гіпотез після того, як стає відомим результат випробувань, в підсумку якого з'явилась подія.

Значення $P(A|B)$ та $P(A|neB)$ підставлені в теорему Байеса дозволяють підрахувати апостеріорну ймовірність результату, тобто ймовірність скореговану в залежності від відповіді користувача на дане запитання:

$$P_{\text{апостеріорна}} = \frac{P_y P}{P_y P + P_n (1 - P)} \quad (2)$$

Значення $P(B|A)$ знаходиться для кожної дієти кожного правила окремо. На виході формується меню страв з коефіцієнтом корисності (від 0 до 1, де 0 – не рекомендується до вживання, 1 – рекомендується до вживання). Страви групуються у список від найбільш корисного до найменш корисного.

Розроблено інформаційну технологію підбору споживацького меню страв у закладах громадського харчування. Досліджено методи створення експертних правил. Алгоритмічно подано систему оцінювання коефіцієнта корисності. Виконано алгоритмічну та програмну реалізацію експертної системи. Застосування технології дає можливість формувати замовлення з урахуванням стану здоров'я клієнта. При формуванні меню враховуються не тільки гастрономічних вподобання, а й дієтологічні вказівки.

Література

1. Elmasri R. Fundamentals of database systems / R. Elmasri, B. Shamkant, T. Navath. – Boston, MA, Addison-Wesley Longman Publishing Co. Inc., 2006. – 543 p.

2. Громадське харчування: організація роботи, документальне оформлення, облік [Електронний ресурс]. – Режим доступу: http://cons.parus.ua/_d.asp?r=03A6Gc88598db68215566c20df2054bb16e0f

3. ІС Управление рестораном и ІС для кафе [Електронний ресурс]. – Режим доступу: <http://iobshepit.com/obzor/>

4. РестАрт 3: Кухня [Електронний ресурс]. – Режим доступу: <https://rarus.ru/1c-restoran/restart-red-3-kukhnya/>

5. Giarratano C. Expert Systems: Principles and Programming / С. Giarratano, G.Riley. – Dallas, TX, Books Worldwide Express, 1998. – 270 p.

6. Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. – Санкт Петербург, «ПИТЕР», 2000. – 375 с.

7. Capinski D. Measure, Integral and Probability / D. Capinski, M. Marek, P. Kopp. – New York, NY, Springer Verlag, 2004. – 170 p.

Інформаційна технологія класифікації зображень на базі SVM

Лесишин М.В., Манзюк Е.А., Скрипник Т.К.

Хмельницький національний університет

На сьогоднішній день існують різні підходи для аналізу та класифікації зображень [1, 2], Деякі одержали реалізацію в комерційних програмних продуктах систем, що входять до складу, автоматизованої мікроскопії.

Використання таких систем дозволяє підвищити точність результатів досліджень, а також провести співставлення між особливостями зображень і результатами інших методів дослідження, зробити результати досліджень більш об'єктивними [3]. Крім того, введення зображень у комп'ютер дозволяє не тільки їх зберігати та сортувати, але та обмінюватися ними з іншими дослідниками, використовуючи для цього сучасні телекомунікаційні засоби.

Класичний підхід, застосовуваний для аналізу та класифікації зображень, широко використовуваний, може бути представлений у вигляді послідовності, що полягає із чотирьох етапів: одержання зображення, попередня обробка, одержання вектора ознак і аналіз отриманих ознак.

У більшості пропонованих на ринку програмних розв'язків для класифікації використовуються формальні ознаки, у першу чергу, морфологічні, що характеризують положення та орієнтацію об'єктів. Поряд з точністю та надійністю результатів такі системи мають і ряд істотних недоліків, і в першу чергу – високу вартість.

У зв'язку із цим представляється актуальною розробка алгоритму класифікації, здатного давати прийнятний результат при роботі зі зразками різної якості та не залежного від апаратної частини аналізатора. Останні роботи з аналізу, класифікації та розпізнаванню зображень показують, що найбільш ефективним у цьому напрямку є використання підходів на основі

змістовного моделювання, коли робота класифікатора або аналізатора будується на основі значень різних характеристик усього зображення, а не формальних ознак досліджуваних об'єктів.

У даній роботі використана класична структура алгоритму класифікації, описана вище. Попередня обробка зображень полягала у виділенні на вихідному зображенні зображень. В основі процедури виділення об'єктів, що цікавлять, лежав алгоритм сегментації кольорових зображень у колірному просторі.

Іншим класом детекторів є морфологічна обробка зображення. Для виявлення подібності використовуються морфологічні операції, такі як перетворення Тор Нат для виявлення світлих і темних частин відповідно. Також можливе використання різних комбінацій морфологічних операцій, таких як відкриття та закриття, ерозія та дилатація з попередньо обраним структуроутворюючим елементом. Даний клас детекторів має у своїй результуючій масці менше неправильних спрацьовувань, чому детектори, засновані на граничній обробці за умови ефективних методів попередньої обробки зображення та постобробки маски. З недоліків впливає неможливість відзначити залежність ефективності роботи від апріорно заданих параметрів.

Ще одним класом способів є методи частотної обробки зображень [4]. Дані методи дозволяють виділяти текстурні особливості на зображенні схожі з базовою функцією.

До останньої групи детекторів слід віднести методи, засновані на навчанні алгоритму. Суть методів навчання полягає у виборі дескрипторів, які найбільше точно характеризують передбачуваний дефект. Ефективність виявлення таких детекторів залежить від якості вибору дескрипторів. Дана група детекторів дозволяє застосовувати кілька методів виявлення дефектів і дає інтегральну оцінку для пікселів або області зображення.

Метою даної роботи є класифікація зображень.

Зображення розглядається як спрощена математична модель, яка являє собою двовимірну дискретну послідовність $Y_{i,j}, i = \overline{1, N}, j = \overline{1, M}$, виду:

$Y_{i,j} = (1 - d_{i,j}) \cdot S_{i,j} + d_{i,j} \cdot c_{i,j}$, де $Y_{i,j}$ - спостережуване зображення, $S_{i,j}$ - оригінальне (неушкоджене) зображення, $d_{i,j}$ - бінарна маска області з перевернутими значеннями (1 - відповідає перевернутим пікселям, 0 - відповідає не перевернутим пікселям), $c_{i,j}$ - перевернуті значення пікселів.

Схема класифікації складається із двох етапів: попередньої обробки з використанням локальних бінарних дескрипторів.

Попередня обробка є важливим кроком при класифікаціях, призначена для придушення шумовий складової, а також усунення не значних дефектів. У даній роботі використовується алгоритм фільтрації.

Дескриптори LBP дозволяють одержати гістограму, яка описує текстурні та структурні особливості зображення. Така гістограма також є інваріантною до повороту та зміни рівня яскравості. Дефект являє собою, як правило, невелику, стосовно фонові області, лінію, яка порушує однорідність текстури тла.

Модифікований LBP оператор може бути записаний у такий спосіб:

$$LBP_{P,K} = \begin{cases} \sum_{p=1}^P f(g_p - g_0) & \text{if } U \leq U_T, \\ P+1 & \text{otherwise} \end{cases}, \quad f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases},$$

де P – число сусідів; g – радіус; U – число переходів між 1 і 0; U_T – граничне значення на кількість переходів.

Таким чином, формується гістограма для локальної області, що представляє собою дескриптор для аналізу локальних областей у зображення.

Для класифікації та співвіднесення кожного пікселя до класу дефектів або до класу об'єктів використовується метод опорних векторів (SVM). Для пояснення роботи методу опорних векторів будемо розглядати завдання класифікації для об'єктів двох класів.

Нехай задані:

– множина X навчальних об'єктів, заданих векторами ознак $X = \{X_1, X_2, \dots, X_n\}$ $X \subset R^d$ (X є підмножиною евклідового простору розмірності d);

– множина Y відповідей для об'єктів навчання: $Y = \{y_1, y_2, \dots, y_n\}$, $y_k \in \{-1, +1\}$ ($k = 1..n$).

Тоді завдання класифікації полягає в побудові такої функції f (класифікатора), яка кожному вектору X_k ($k = 1..n$) співставляється правильна відповідь y_k .

У методі SVM у якості функції f обрана площина, відстані до якої найближчих векторів обох класів рівні. Найближчі точки-вектори називаються опорними. При цьому для всіх об'єктів одного класу повинна виконуватися нерівність $f(X_k) > 0$, а для всіх об'єктів іншого класу – нерівність $f(X_k) < 0$.

Рівняння поділяючої площини має такий вигляд: $w_1x_1 + w_2x_2 + \dots + w_dx_d + w_0 = 0$, де d – розмірність підпростору ознак; $W = (w_1, w_2, \dots, w_d)$ – напрямний вектор, w_0 – скалярний поріг.

У методі опорних векторів виділяють два етапи: етап навчання та етап розпізнавання. На першому етапі з множини навчальних прикладів відбираються опорні вектори, на основі яких будується поділяюча площина. Етап розпізнавання полягає в тому, що на вхід отриманого класифікатора

подається приклад X , про класову приналежність якого нічого не відомо. Класифікатор повинен дати відповідь, до якої класу відноситься вектор X .

Для побудови поділяючої гіперплощини використовується радіальна базисна функція:

$$k(x, x') = \exp(-\gamma \|x - x'\|^2) \text{ при } \gamma > 0,$$

де x – поточний вектор, x' – центральний вектор, γ – параметр, що нормалізує.

Попередня обробка допомагає зменшити вплив шумовий складової, а також усунути невеликі дефекти.

Отже, в роботі представлений автоматизований алгоритм класифікації на основі модифікованого оператора локальних бінарних околиць. Для класифікації дескрипторів і поділу на класи використаний метод опорних векторів. Приклади, представлені в роботі, демонструють ефективність алгоритму при класифікації на складно текстурних зображеннях.

Література

1. Lin Jin, Liu Yayu. Potholes Detection Based on SVM in the Pavement Distress Image, Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science, pp. 544 – 547, 2010.
2. Методы компьютерной обработки изображений./Под редакцией В.А. Сойфера – М.: Физматлит, 2003.-784с.
3. Barni M., Bartolini F., Cappellini V. Image processing for virtual restoration of artworks. IEEE Multimedia, vol. 7, no. 2, pp. 34-37, 2000.
4. Zhou J., Huang P., and Chiang F. Wavelet-based pavement distress classification. TRR: Journal of the Transportation Research Board, pp. 89–98, 2005.

Метод формального опису елементів моделей автоматизованого формування тестових завдань

Мазурець О.В., Ковальчук О.В., Слободзян В.О., Білоус Г.А.
Хмельницький національний університет

Одним із основних способів контролю знань в навчальних інформаційних системах є комп'ютерне тестування, яке крім контрольної функції застосовується для навчання, тренінгу, розвитку здібностей. Інформаційні технології дають можливість суттєво зменшити трудові затрати на створення тестових завдань з можливістю їх постійного оновлення, що формує актуальний напрямок наукових досліджень.

Інформаційна технологія автоматизованого формування тестових завдань на основі контенту навчальних матеріалів [1] дозволяє на основі вхідних даних у вигляді контенту файлу формату .docx навчальних матеріалів або його визначеної частини автоматизовано формувати вихідні

дані у вигляді файлу з тестами для імпорту у віртуальне навчальне середовище. Розв'язок задачі автоматизації формування тестових завдань містить ряд послідовних етапів перетворення інформації, до яких відносяться визначення семантичних термінів у контенті навчальних матеріалів [2], формування структури навчальних матеріалів [3], безпосередньо автоматизоване формування тестових завдань [1] й перенесення результатів у область взаємодії з кінцевим користувачем – підсистему тестування середовища Moodle [4].

Характерною рисою автоматизованого формування тестових завдань за даною інформаційною технологією є використання набору моделей тестових завдань для перетворення фрагментів контенту у тестові завдання.

Модель тестового завдання являє собою структуру, що складається з наступних функціональних елементів:

- маска для фрагменту тексту з поняттям, що призначена для ідентифікації фрагментів контенту із заданим терміном, до яких воно може бути застосовано;

- параметри моделі тестового завдання, що визначають особливості й ефективність його застосування: тип запитання, що може бути створено за цією моделлю; базова оцінка очікуваної якості застосування моделі у діапазоні (0;1), що одержується методом експертної оцінки результатів; набір правил корекції базової оцінки цього правила конвертації, що в залежності від вказаних особливостей фрагменту можуть знизити базову оцінку, при активації виступаючи множниками у діапазоні (0;1);

- маска для формування тестового завдання, що містить алгоритм перетворення даного фрагменту у елементи тестового завдання.

Тобто модель у даному випадку представляє собою структуру, в яку входять: набір з'єднувачів (слів або символів які вирізняють певний фрагмент тексту як релевантний відносно терміну); правило конвертації фрагментів з тексту у тестове завдання, та дані про словоформу (відмінок/рід). На вхід моделі завжди дається певна зв'язка [термін – слово маркер – теза], за якою здійснюється формування тестового завдання та підбір варіантів відповідей.

При використанні даного підходу спершу встановлюються вимоги до набору тестових завдань, визначаються актуальні моделі тестових завдань й на основі найбільш прийнятних із них формується набір тестових завдань. В результаті за кожною з обраних моделей тестових завдань формується одне тестове запитання у форматі .gift, що може бути конвертоване у середовище Moodle. Кінцевий результат зберігається в одному файлі тесту.

Таким чином, при формуванні тестових завдань різних типів використовуються моделі, наділені параметрами та критеріями, характерними лише для конкретного типу питання та будови терміну. В межах розглянутої інформаційної технології, з метою формування моделей

тестових завдань є необхідною розробка набору тегів для формального опису елементів правил конвертації.

Метою роботи є розробка множини тегів як методу формального опису елементів моделей автоматизованого формування тестових завдань. Це дозволить проводити як зручне й ефективне створення нових чи редагування існуючих моделей формування тестових завдань, так і компактне зберігання розроблених моделей для подальшого використання.

Відповідно до розглянутих складових моделей формування тестових завдань, виділяються теги для ідентифікації елементів контенту та теги для формування тестів. Теги для ідентифікації елементів контенту використовуються в масках ідентифікації й наведені в таблиці 1.

Таблиця 1 – Теги для ідентифікації елементів контенту

Тег	Опис
[TermGroup]	Фрагмент тексту, який являє собою термін, що складається з набору слів
[ThesisGroup]	Фрагмент тексту (теза), який дає визначення терміну
[RandomTermGroup]	Випадково обраний інший термін
[RandomThesisGroup]	Випадково обране визначення іншого терміну
[Connector]	Слово або символ із тексту, що поєднує термін з тезою (–, – це, є, називається, тощо)
[BeginSentence]	Фрагмент тексту від початку речення до TermGroup або ThesisGroup (може бути null)
[Inflexion]	Тег повертає нормальну форму наступного елементу
[Caps]	Переведення першої букви до верхнього регістру
[ReCaps]	Переведення першої букви до нижнього регістру

Нижче наведений приклад для ідентифікації двох моделей з істинним твердженням, базової (Basic) та реверсивної (Reversed):

[Caps][TermGroup][connector][ThesisGroup]. {TRUE}

[Caps][ThesisGroup][connector][TermGroup]. {TRUE}

Теги для формування тестів використовуються в масках формування й наведені у таблиці 2.

Таблиця 2 – Теги для формування тестів

Тег	Опис
{Header}	Візуальний визначник для тексту питання
{Answer}	Візуальний визначник для варіанту відповіді
{FALSE}	Вказівка на неправильний варіант відповіді
{TRUE}	Вказівка на правильний варіант відповіді

Нижче зображений приклад моделі формування тестового завдання одиничного вибору, де в якості тексту питання (Header) вказана теза, а у

варіантах відповідей (Answer 1 – N) назви термінів:

```
Header -> [Caps][ThesisGroup][“-”, - це”]
Answer 1 -> [TermGroup] {TRUE}
Answer 2 -> [RandomTermGroup] {FALSE}
Answer 3 -> [RandomTermGroup] {FALSE}
Answer N -> [RandomTermGroup] {FALSE}
```

Наведені теги можна віднести до елементів псевдокоду, або змінних, які під час роботи алгоритму мають різні значення, а іноді можуть не мають жодних (null). У випадках, коли тег може повернути порожнє значення, він вказується з додаванням знака дорівнює і нуля, наступним чином: [Tag = 0].

Наведений підхід дозволяє відкрите програмування алгоритму роботи різноманітних моделей генерування тестових завдань, що визначає множину формалізованих таким чином моделей як базу знань відповідної інформаційної системи для автоматизованого формування тестових завдань за навчальними матеріалами.

Для прикладу наведено процес формування тестового завдання типу «із введенням тексту» з використанням однієї з моделей формування тестових завдань для відповідного типу завдань.

Моделі формування завдань із введенням тексту полягають у формулюванні одного твердження з відсутнім ключовим словом та формуванні можливих варіантів відповідей, які не пропонуються користувачеві, а використовуються лише для перевірки введеного тексту. Відповідно, можлива модель, мета якої – забезпечити максимальну кількість правильних відповідей для коректної перевірки тексту, введеного користувачем.

При формуванні маски ідентифікації, за основу твердження береться певна пара [термін – теза], причому важливо, щоб термін був у початковій формі (називному відмінку однини), щоб виключити можливість ігнорування правильних відповідей через несхожість у закінченнях або словоформах. Далі відбувається формування набору правильних відповідей. Необхідним і достатнім є хоча б один варіант відповіді, якщо він повністю покриває можливі форми вживання даного терміну. Множина варіантів відповідей формується на підставі параметрів моделі.

При формуванні правильних відповідей необхідно враховувати: абрєвіатуру терміну; скорочення; слова іншомовного походження (якщо є у тексті); відмінок/рід терміну.

Оскільки варіанти відповідей завжди сховані, функції цієї моделі не включають у себе підбір хибних термінів, а навпаки – модель має гарантувати, що всі можливі форми вживання цього поняття будуть розпізнані як правильна відповідь.

Отже, для сформованого на рисунку 1 тестового питання типу «Коротка відповідь» було використано термін «Вимірювання» та відповідну модель формування тестових завдань із відкритою відповіддю. Термін взятий з фрагменту тексту: «Вимірювання – це множина однотипних даних, що

утворюють одну з граней гіперкуба» з теми «Моделі даних» навчального матеріалу «Організація баз даних та знань».

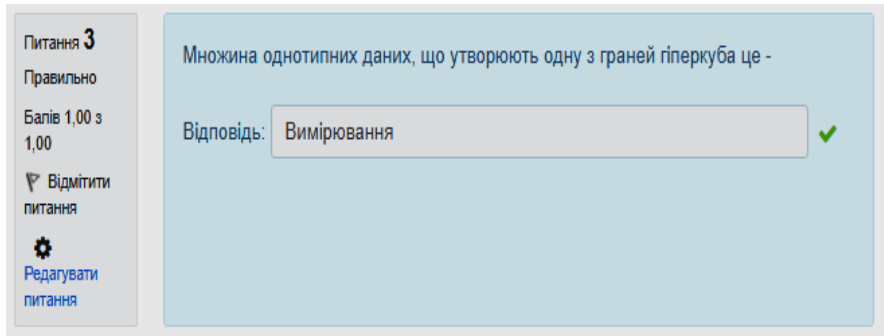


Рисунок 1 – Приклад використання у Moodle сформованого тестового завдання типу «Коротка відповідь»

Необхідно відмітити, що правильними відповідями є обидві з наведених, а для повної відповіді на питання достатньо вказати одну з них.

В даному випадку актуальний фрагмент тексту був знайдений наступною маскою ідентифікації:

```
[BeginSentence][ReCaps][TermGroup][Connector][ThesisGroup].
```

А генерація тестового завдання за актуалізованою моделлю відбулася шляхом використання відповідної маски формування тестового завдання із таким кодом:

```
Header -> [Caps][_____][connector][ThesisGroup]:  
Answer 1 -> [Abbreviation] {TRUE}  
Answer 2 -> [TermGroup1] {TRUE}  
Answer N -> [TermGroupN] {TRUE}
```

Відповідний сформований фрагмент GIFT-файлу, який сформований за даною моделлю формування тестового завдання типу «із введенням тексту» із відкритою відповіддю, наступний:

```
// question  
Множина однотипних даних, що утворюють одну з граней  
гіперкуба це -  
{  
=%100% Вимірювання#  
=%100% Dimension#  
}
```

Отже, розроблені теги для моделей формування тестових завдань дозволяють формально описати процес формування тестових завдань із рахуванням всіх особливостей та параметрів, й забезпечити автоматизацію імпорту доступних тестових завдань у середовищі Moodle. Тег формального

опису моделей є елементом псевдокоду, який призначений для формального опису структури моделі, її вхідних та вихідних параметрів.

За умови достатньої відповідності семантичним та структурним вимогам і коректного співвідношення між обсягом контенту навчального матеріалу та параметрів генерації набору тестових завдань, одержується репрезентативний тест, що може бути використаний як безпосередньо для тестування, так і як сировина для подальшої роботи розробника тестів.

Література

1. Бармак О. В., Мазурець О. В., Кліменко В. І. Інформаційна технологія автоматизованого формування тестових завдань / О. В. Бармак, О. В. Мазурець, В. І. Кліменко // Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2017, №5. – С.93-103.

2. Мазурець О. В. Інформаційна технологія автоматизованого визначення семантичних термінів в елементах навчальних матеріалів / О. В. Мазурець // Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2018, №3. – С.223-230

3. Мазурець О. В. Онтологічний підхід до побудови семантичної моделі навчальних матеріалів / О. В. Мазурець // Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2017, №6. – С.223-229.

4. Moodle – Open-source learning platform. – [Електронний ресурс]. – Режим доступу: <https://moodle.org/>

Система підбору методом спільної фільтрації продукції інтернет-магазину літератури

Мартинюк Б.І., Лищук О.А.

Хмельницький національний університет

З розвитком економічних відносин у суспільстві, виникла потреба швидко і ефективно збувати продукцію. А в людей, що купують товари виникла потреба максимально раціонально використовувати власний час. Саме тому інтенсивний розвиток веб-технологій зумів задовільнити всі сторони даних суспільних відносин. Зокрема почали з'являтися інтернет-магазини, які з кожним роком стають більш функціональними і традиційні магазини не здатні з ними конкурувати. Адже в електронній комерції є такі функції, які не здатні використовувати традиційний метод торгівлі. А саме йдеться про потужний функціонал швидкого підбору товару, де немає людського фактору, що значно пришвидшує пошук необхідного товару для

користувача.

В електронній комерції ці функції називаються фільтрація продукції і рекомендаційні системи. Дані функції реалізувати можна за допомогою різних математичних і соціологічних алгоритмів, які дуже рідко помиляються, якщо мова йде про рекомендацію товару користувачу на сайті. Також важливу роль відіграє фільтрація на сайті, яка відшукує товар за заданими параметрами секунди при тому, що на сайті можуть бути десятки тисяч різних продуктів.

Ефективна фільтрація - це найшвидший спосіб отримати споживачів від домашньої сторінки до сторінки детальної інформації про продукт. Фільтри електронної комерції продукту, що використовуються неправильно, можуть відлякати потенційних користувачів інтернет ресурсу. Тому було сформовано 31 розділ, який допоможе ефективно оцінити фільтрацію на сайті.

Рекомендаційна система[1] — підклас системи фільтрації інформації, яка будує рейтинговий перелік об'єктів (фільми, музика, книги, новини, веб-сайти), яким користувач може надати перевагу. Для цього використовується інформація з профілю користувача. Існують дві основні стратегії створення рекомендаційних систем[2]: фільтрація вмісту і колаборативна фільтрація.

При колаборативній фільтрації використовується інформація про поведінку користувачів у минулому — наприклад, інформація про придбання або оцінки. В цьому разі не має значення, з якими типами об'єктів ведеться робота, але при цьому можна брати до уваги неявні характеристики, які складно було б врахувати при створенні профілю. Основна проблема цього типу рекомендаційних систем — «холодний старт»: відсутність даних про користувачів чи об'єкти, які нещодавно з'явилися у системі.

Є різні методики збору даних. У процесі роботи рекомендаційні системи збирають дані про користувачів, використовуючи поєднання явних і неявних методів.

Приклади явного збору даних:

- користувач оцінює запропонований об'єкт за диференційованою шкалою;
- користувач ранжує групу об'єктів від найкращого до найгіршого;
- користувач вибирає кращий з двох запропонованих об'єктів;
- користувачу пропонують створити список його улюблених об'єктів.

Приклади неявного збору даних:

- спостереження за тим, що користувач оглядає в інтернет-магазинах або базах даних іншого типу;
- ведення записів про поведінку користувача онлайн;
- відстеження вмісту комп'ютера користувача;

Існують різні алгоритми, які допомагають полегшити процес фільтрації. Одним з таких алгоритмів є фільтрування на основі вмісту. Цей

алгоритм рекомендує продукти, подібні до тих, які користувачу сподобалися раніше. Вони зберігають всю інформацію, що стосується кожного користувача у векторній формі. Цей вектор містить минуле поведінки користувача, тобто фільми, які сподобалося / не подобалося користувачеві та надані ними рейтинги. Цей вектор відомий як профільний вектор.

Алгоритм фільтрування на основі вмісту знаходить косинус кута між вектором профілю та елементом вектора, тобто косинусною подібністю. Припустимо, що A - профільний вектор, а B - вектор елемента, тоді подібність між ними може бути розрахована як:

$$sim(A, B) = \cos \theta = \frac{A \cdot B}{\|A\| \|B\|}$$

Основний недолік цього алгоритму полягає в тому, що він обмежується рекомендацією елементів одного типу. Він ніколи не рекомендує продукти, які користувач раніше не купував або не любив. Тож якщо користувач переглянув або відмітив як улюблене книги про фентезі, система буде рекомендувати лише книги про фентезі. Це дуже вузький спосіб побудови рекомендаційної системи.

Тому розглянемо наступний метод, який називається спільна фільтрація. Це один з найпоширеніших алгоритмів у галузі, оскільки він не залежить від будь-якої додаткової інформації. Цей алгоритм спочатку знаходить оцінку подібності між користувачами. Виходячи з цієї оцінки подібності, він потім вибирає найбільш подібних користувачів і рекомендує продукти, які подібні користувачі сподобались або купували раніше.

Розглянемо детальніше спільну фільтрацію типу користувач-користувач. Цей алгоритм спочатку знаходить оцінку подібності між користувачами. Виходячи з цієї оцінки подібності, він потім вибирає найбільш подібних користувачів і рекомендує продукти, які подібні користувачі сподобались або купували раніше.

Передбачення елемента для користувача U обчислюється шляхом обчислення зваженої суми оцінок користувачів, дані інших користувачів до елемента i .

Прогноз $P_{u,i}$ задається наступним чином:

$$P_{u,i} = \frac{\sum_v (r_{v,i} * s_{u,v})}{\sum_v s_{u,v}}$$

де $P_{u,i}$ - прогноз товару; $r_{v,i}$ - це рейтинг, заданий користувачем v у фільмі i , $s_{u,v}$ - подібність між користувачами.

Цей алгоритм вимагає багато часу, оскільки він включає в себе обчислення подібності для кожного користувача, а потім обчислення прогнозу для кожної оцінки подібності. Один із способів вирішення цієї проблеми - вибрати лише декілька користувачів (сусідів) замість того, щоб робити прогнози, тобто замість того, щоб робити прогнози для всіх значень подібності, ми вибираємо лише деякі значення подібності.

Тому розглянемо детальніше спільну фільтрацію, яка працює за принципом товар-товар. У цьому алгоритмі ми обчислюємо подібність між кожною парою елементів. Цей алгоритм працює подібно до спільної фільтрації користувач-користувач з лише невеликою зміною - замість того, щоб взяти зважену суму рейтингів "сусідів-користувачів", ми приймаємо зважену суму рейтингів "сусідів за об'єктами". Прогноз визначається:

$$P_{u,i} = \frac{\sum_N (s_{i,N} * R_{u,N})}{\sum_N (|s_{i,N}|)}$$

Також виникає проблема для користувача, який щойно зареєструвався і не зробив ще жодної покупки або таж проблема для нового товару. Для них не можна сформувати матрицю. Це називається «Cold Start» [3]. Може бути два типи холодного старту

1. Відвідувач Cold Start.
2. Продукт Cold Start.

Відвідувач Cold Start означає, що в набір даних вводиться новий користувач. Оскільки немає історії цього користувача, система не знає переваг цього користувача. Важко рекомендувати продукти цьому користувачеві. Одним з основних підходів вирішення даної проблеми може бути застосування популярної стратегії, тобто рекомендувати найпопулярніші продукти. Це можна визначити тим, що останнім часом було популярним загалом чи регіонально. Як тільки будуть відомі вподобання користувача, рекомендувати товари буде легше.

Провівши відповідні дослідження було реалізовано програмний продукт на основі даних алгоритмів. Було також враховано недоліки схожих програмних продуктів. Результатом було реалізовано інтернет-магазин літератури на CMS WordPress. Для даного сайту було розроблено розширення у вигляді рекомендаційної системи на основі вище описаних алгоритмів. Також була розроблена система фільтрації продукту на сайті за параметрами, які може задати користувач сайту. Оскільки на дану систему мало схожих програмних продуктів, то дана розробка є актуальною.

Література

1. Melville P., Mooney R., Nagarajan R. Content-Boosted Collaborative Filtering for Improved Recommendations// University of Texas, USA : Матеріали конф. / AAAI-02, Austin, TX, USA, 2002. — 2002. — P. 187-192.
2. Matrix Factorization Techniques for Recommender Systems. Computer (IEEE) 42 (8): 30–37. – Режим доступу: Koren, Y.; Bell, R.; Volinsky, C. (07 August 2009).
3. Lika, Blerina; Kolomvatsos, Kostas; Hadjiefthymiades, Stathes (March 2014). "Facing the cold start problem in recommender systems". Expert Systems with Applications.

Інформаційна система забезпечення контролю фінансів ведення малого бізнесу

Медцовський О. В.

Хмельницький національний університет

В розвитку економіки України важлива роль відведена малому бізнесу, особливо це стосується регіонального розвитку. Малий бізнес може включати в себе індивідуальну трудову діяльність громадян, що мають статус юридичної особи або не мають його. До малого бізнесу відносяться індивідуальні, приватні, сімейні підприємства, фермерські господарства тощо. Незважаючи, на те, що підприємства малого бізнесу можуть працювати в найрізноманітніших сферах економіки для всі них досить гостро стоїть проблема контролю фінансів [1]. Особливо це стосується окремих громадян, що ведуть індивідуальну трудову діяльність.

Для ефективного розподілу своїх доходів та витрат, людина має володіти мінімальними навиками фінансової грамотності. Постійний облік доходів та витрат, використання інструментів обліку та аналізу особистих фінансів дозволять забезпечити фінансову стабільність, закладуть основи добробуту та дозволять акумулювати фінансові ресурси для подальшого розвитку.

Існує цілий ряд готових програмних рішень, щодо забезпечення контролю фінансів. Це такі як «Домашня бухгалтерія», Moneytracker, Personal Finances, My Budget, Icontrolmymoney, Icash, DomEconom, AbilityCash, Buddi, Все Pro Гроші, Свої деньги тощо [2]. Здавалося би, при такому переліку готових рішень немає потреби у розробці нової інформаційної системи забезпечення контролю фінансів ведення малого бізнесу. Але всі ці готові програмні рішення маю одну суттєву ваду – це додатки, які можуть працювати на комп'ютері, але не можуть працювати через мережу Інтернет і веб-інтерфейс. А в час інтенсивного розвитку інформаційних технологій, коли майже все можна робити через мережу Інтернет, є нагальна потреба у інформаційній системі контролю фінансів ведення малого бізнесу з веб-інтерфейсом, коли вести контроль за своїми фінансами можна би було у режимі реального часу. Щоб у будь-який час можна було б отримати достовірну інформацію щодо доходів та витрат, та лімітів фінансових ресурсів, які є доступними на певний час. Це підвищить обґрунтованість прийняття рішень в сфері особистих фінансів, збільшить горизонт планування та закладе основи фінансової стабільності малого бізнесу.

Метою роботи є розробка інформаційної системи забезпечення контролю фінансів малого бізнесу, яка буде забезпечувати базові потреби підприємства в обліку фінансів. Інформаційна система забезпечення контролю фінансів малого бізнесу призначена для громадян, що ведуть індивідуальну трудову діяльність та окремих користувачів, що мають потребу у контролі особистих фінансів.

Враховуючи, вище викладене, сформулюємо функціональні та

нефункціональні вимоги до інформаційної системи.

Функціональні вимоги:

- зручна форма підрахунку доходів та витрат,
- зручна форма ведення статистики у вигляді діаграм,
- можливість розподілу всіх статей доходів та витрат за категоріями та групами;
- можливість чітко планувати всі статті доходів та витрат;
- можливість встановлення лімітів, які не бажано перевищувати;
- можливість конвертації грошових коштів в основні валюти, такі як долар та євро по курсу он-лайн банкінгу.

Нефункціональні вимоги:

- створення простої бази даних;
- можливість реєстрації та авторизації;
- захист від несанкціонованого доступу до інформаційної системи шляхом встановлення паролів;
- можливість використання різних браузерів для роботи з інформаційною системою;
- можливість використання різних гаджетів для он-лайн доступу до інформаційної системи.

Після визначення вимог до інформаційної системи була спроектована модель системи, що представлена на рисунку 1.

Коротко опишемо роботу спроектованої інформаційної системи забезпечення контролю фінансів ведення малого бізнесу.



Рисунок 1 – Модель інформаційної системи забезпечення контролю фінансів ведення малого бізнесу

Модуль користувачів призначений для зберігання даних про зареєстрованих користувачів інформаційної системи та їх облікових даних. Адміністративний модуль призначений для адміністратора інформаційної системи, який буде формувати перелік користувачів, розмежовувати права доступу тощо. Модуль контенту буде зберігати данні про події (доходи та витрати конкретного користувача, детальну інформацію про кожну подію). Модуль пошуку призначений для організації пошуку по історії подій конкретного користувача. Модуль звітів призначений для виведення в графічній формі інформації по контролю фінансів, загальних сум по кожній категорії витрат, ліміти по кожній категорії та інформацію про перевищення лімітів.

При вході на веб-сторінку користувач повинен авторизуватися. Якщо користувач вперше заходить на сторінку, то є можливість пройти реєстрацію. Після авторизації користувачу буде доступний весь функціонал інформаційної системи. До пунктів меню верхнього рівня відносяться такі пункти: «Рахунок», «Історія», «Планування», «Запис». В меню «Історія» виводиться історія подій, які були додані користувачем. Є можливість подивитися детальну інформацію про кожну подію, здійснювати пошук по різним параметрам та сортування даних. В меню «Планування» можна виводити графічну інформацію щодо стану фінансів, побачити загальні суми по кожній категорії та отримати інформацію щодо ліміта з кожної категорії. В меню «Запис» користувач має можливість додавати категорії витрат, встановлювати їх ліміти, додавати події до різних категорій, вибираючи їм тип. Також можна редагувати подію з вибраної категорії. В меню «Рахунок» виводиться конвертація фінансових ресурсів користувача в такі валюти як долар і євро в режимі он-лайн банкінгу.

Отже, проведені дослідження було направлено на виявлення суті та особливостей інформаційної системи забезпечення контролю фінансів малого бізнесу, призначеної в першу чергу для громадян, що займаються індивідуальною трудовою діяльністю. Були вивчені потреби користувачів, здійснений збір вимог, розроблені специфікації вимог. У відповідності до специфікацій вимог користувачів була розроблена модель інформаційної системи. Здійснена програмна реалізація спроектованої інформаційної системи. Запропонована модель інформаційної системи забезпечення контролю фінансів малого бізнесу представляє собою простий та надійний інструмент контролю особистих фінансів.

Подальші дослідження будуть спрямовані на вдосконалення технологій захисту персональних даних користувачів та технологій доступу з різних видів гаджетів.

Література

1. Сиротинська А.П. Інформаційні системи підприємств малого бізнесу: Навч. посібник / А.П. Сиротинська, І.Д.Лазаришина. – Київ : Цунтр учбової літератури, 2008. – 264 с.

2. Домашняя бухгалтерия. Обзор основных предложений на рынке.– Все Про гроші [Електронний ресурс]. – Режим доступу: <http://vseprogroshi.com.ua/ru/domashnyaya-buxgalteriya-obzor.html>

Інформаційна система генерування автоматизованих рекомендацій на базі активності користувача

Мельник М.О., Плетюк М.М., Коломієць Н.О.

Хмельницький національний університет

Інформаційні системи, як інформація і інформаційні технології, існували з моменту появи суспільства, оскільки на будь-якій стадії його розвитку є потреба в управлінні. Організація управління виробництвом (послугами) засновано на взаємодії керуючої (менеджер) і керованої (виробничо-технологічної) частин. Фактичний стан керованого (виробничо-технологічного) процесу постійно порівнюється із необхідним станом за вимогами результату управління. Таке порівняння здійснюється за використанням інформаційних технологій, тому що для управління потрібна систематизована, заздалегідь підготовлена інформація про необхідний і фактичний стан керованої частини. Отже, вироблення інформації - місія інформаційних систем.

Господарство країни, як система, складається з виробничо-технологічної та інформаційної систем. Перша частина - об'єкт управління, друга - засіб для управління. Обидві системи перебувають у постійній взаємодії і взаємозалежності, які забезпечуються системами комунікацій. Реалізація функцій систем і їх корегування здійснюється через систему управління. А весь цей складний механізм працює задля головної мети - зростання ефективності господарської діяльності.

Сьогодні, у вік інформації та комп'ютеризації інформаційні ресурси є такими самими ресурсами, як і трудові, матеріальні та енергетичні. Отож ми можемо казати про інформаційну економіку, що ґрунтується на інформації, та інформаційну сферу – керівників різних рівнів, учених, спеціалістів і службовців. Управління економікою можна порівняти з управлінням живим організмом, де головну роль відіграє нервова система.

В економіці роль нервової системи відіграють інформаційні потоки. Порушення в потоках інформації призводять до перебоїв як в роботі конкретного підприємства, так і всього господарства, а без достовірної, повної та своєчасної інформації неможливо керувати будь-яким виробництвом.

Обчислювальна техніка все ширше використовується як в управлінні виробничими процесами, так і всією економікою. Але матеріальні витрати на зберігання, передавання та переробку інформації вже зараз перевищують аналогічні витрати в світі на енергетику.

Рівень цивілізації можна визначати за рівнем телефонізації та комп'ютеризації суспільства. Так, у США на одну тисячу чоловік припадає 900 телефонів і 140 комп'ютерів, а в нас – відповідно 90 і 2.

Комп'ютеризація суспільства – це не лише техніка, а й люди. Нова комп'ютерна ідеологія має пронизувати всі ланки підготовки спеціалістів і керівників. Безграмотний в комп'ютерному відношенні керівник не може й сподіватися на прогрес у своїй роботі.



Рисунок 1 – Схема формування інформаційної технології

Нову економічну думку та механізм управління неможливо втілити в життя без комп'ютеризації, без нових інформаційних технологій, бо вони мають базуватися на точній, достовірній та своєчасній інформації.

В Україні створене Національне агенство по питаннях інформатизації при Президентіві України та затверджені закони України “Про Національну програму інформатизації (74/98-ВР від 04.02.98) і “Про концепцію Національної програми інформатизації (75/98-ВР від 04.02.98).

Тому інформаційна технологія – цілеспрямована організована сукупність інформаційних процесів з використанням засобів обчислювальної техніки, що забезпечують високу швидкість обробки даних, швидкий пошук інформації, розосередження даних, доступ до джерел інформації незалежно від місця їх розташування.

Мета створення інформаційних систем – у гранично короткі терміни створити систему обробки даних, яка має задані споживчі якості. До них належать: функціональна повнота, своєчасність, функціональна надійність, адаптивна надійність, економічна ефективність.

Функціональна повнота – це властивість інформаційної системи, яка характеризує рівень автоматизації управлінських робіт.

Коефіцієнт функціональної повноти

$$K_f = \frac{\Pi_a}{\Pi_o},$$

де Π_a – показники, отримувані автоматизовано; Π_o – загальна кількість показників.

Таблиця 1 – Різновиди та призначення АСУ

Вид автоматизованої системи управління	Призначення автоматизованої системи управління
Територіальні	Управління адміністративно-територіальним регіоном (областю, містом, районом)
Галузеві	Управління підвідомчими підприємствами та організаціями
Корпоративні	Вищення задач управління виробничо-господарською діяльністю підприємства
АСУ ТП	Укравління технологічними процесами на підприємстві
Інтегровані (ІАСУ)	Багаторівневі ієрархічні системи, які забезпечують комплексну автоматизацію управління на всіх його рівнях

Своєчасність – це властивість інформаційної системи, яка характеризує можливість отримання апаратом керівництва необхідної інформації.

Коефіцієнт своєчасності

$$K_c = \frac{\nabla \Pi_a - \Pi_a}{\Pi_a},$$

де Π_a – кількість показників, отриманих із затримкою щодо планового терміну подання.

Функціональна надійність – це властивість інформаційної системи виконувати свої функції з обробки даних. Це сукупність надійностей програмного, інформаційного та технічного забезпечення.

Адаптивна надійність – це властивість інформаційної системи виконувати свої функції, якщо вони змінюються в межах умов, зумовлених розвитком системи керування об'єкта впродовж заданого проміжку часу.

Економічна ефективність інформаційної системи виявляється в покращенні економічних результатів функціонування об'єкта в результаті впровадження інформаційної системи.

Література

1. Інформаційні системи [Електронний ресурс]. – Режим доступу: <http://www.topaz.ho.ua/Is/1.html>
2. Сиротинська А.П. Інформаційні системи підприємств малого бізнесу: Навч. посібник / А.П. Сиротинська, І.Д.Лазаришина. – Київ : Цунтр учбової літератури, 2008. – 264 с.

Особливості впровадження сучасних експертних систем

Мосьондз В.О., Петровський С.С.

Хмельницький національний університет

Сучасні темпи технологічного та інформаційного розвитку суспільства визначають постановку задачі організації ефективного управління освітніми закладами з можливістю отримання інформації в реальному часі для прийняття науково обгрунтованих управлінських рішень. Для якісного проведення навчально-виховного процесу і використання в ньому можливостей дистанційної освіти також необхідні експертні системи управління.

Метою роботи є детальне ознайомлення та дослідження експертних систем, їхньою будовою та перспективами, що дасть можливість розробки інформаційної технології впровадження експертної системи управління освітніми закладами для управління навчально-виховним процесом.

Розглянемо поняття та перспективи впровадження експертних систем. Експертна система (ЕС, expert system) - комп'ютерна програма, здатна частково замінити фахівця-експерта у вирішенні проблемної ситуації. В інформатиці експертні системи розглядаються спільно з базами знань як моделі поведінки експертів у певній галузі знань з використанням процедур логічного висновку і прийняття рішень, а бази знань як сукупність фактів і правил логічного висновку в обраній предметній галузі діяльності. Схожі дії виконує програма-майстер (wizard).

Майстри застосовуються як в системних програмах, так і в прикладних для інтерактивного спілкування з користувачем (наприклад, при установці ПЗ). Головна відмінність майстрів від ЕС - відсутність бази знань; всі дії жорстко запрограмовані. Це просто набір форм для заповнення користувачем. Інші подібні програми - пошукові або довідкові (енциклопедичні) системи. За запитом користувача вони надають найбільш підходящі (релевантні) розділи бази статей (уявлення про об'єкти областей знань, їх віртуальну модель).

Дослідники в області ЕС для назви своєї дисципліни часто використовують також термін «інженерія знань», введений Е. Фейгенбаум як «привнесення принципів та інструментарію досліджень з області штучного інтелекту у вирішення важких прикладних проблем, які потребують знань експертів». Програмні засоби (ПС), що базуються на технології експертних систем, або інженерії знань (надалі будемо використовувати їх як синоніми), набули значного поширення у світі. Важливість експертних систем полягає в наступному [2,3]:

1. Технологія експертних систем істотно розширює коло практично значущих завдань, що вирішуються на комп'ютерах, вирішення яких приносить значний економічний ефект;

2. Технологія ЕС є найважливішим засобом у вирішенні глобальних проблем традиційного програмування: тривалість і, отже, висока вартість

розробки великих програм;

3. Висока вартість супроводу складних систем, яка часто в кілька разів перевершує вартість їх розробки; низький рівень повторної використовуваних програм та ін.;

4. Об'єднання технології ЕС з технологією традиційного програмування додає нові якості до програмних продуктів за рахунок: забезпечення динамічної модифікації додатків користувачем, а не програмістом; більшої «прозорості» додатка: кращої графіки; інтерфейсу і взаємодії.

На думку фахівців, в перспективі ЕС знайдуть наступне застосування:

1. ЕС будуть відігравати провідну роль у всіх фазах проектування, розробки, виробництва, освіти розподілу, продажу, підтримки і надання послуг;

2. Технологія ЕС, що отримала комерційне поширення, забезпечить революційний прорив в інтеграції додатків з готових інтелектуально-взаємодіючих модулів.

ЕС призначені для так званих неформалізованих задач, тобто ЕС не відкидають і не замінюють традиційного підходу до розробки програм, орієнтованого на рішення формалізованих завдань [4].

Неформалізовані задачі зазвичай володіють наступними особливостями:

1. Помилковістю, неоднозначністю, неповнотою і суперечливістю вихідних даних;

2. Помилковістю, неоднозначністю, неповнотою і суперечливістю знань про проблемну області і розв'язуваної задачі;

3. Великою розмірністю простору рішення, тобто перебір у пошуку рішення дуже великий;

4. Динамічно змінюються даними і знаннями.

Для функціонування ЕС можна поділити на 2 режиму роботи:

1. Режим введення знань - в цьому режимі експерт з допомогою інженера по знаннях за допомогою редактора бази знань вводить відомі йому дані про предметну область в базу знань ЕС.

2. Режим консультації - користувач веде діалог з ЕС, повідомляючи їй відомості про поточну завдання і отримуючи рекомендації ЕС. Наприклад, на основі відомостей про фізичний стан хворого ЕС ставить діагноз у вигляді переліку захворювань, найбільш ймовірних при даних симптомах.

Технологію побудови експертних систем називають інженерією знань. Цей процес вимагає специфічної форми взаємодії творця експертної системи, якого називають інженером знань, і одного або кількох експертів в деякій предметній області. Інженер знань «витає» з експертів процедури, стратегії, емпіричні правила, які вони використовують при вирішенні завдань, і вбудовує ці знання в експертну систему. У результаті з'являється система вирішувача завдання багато в чому так само, як людина – експерт (рисунок 1).

Структура ЕС складається [1]:

1. Інтерфейс користувача.

2. Користувач.
3. Інтелектуальний редактор бази знань.
4. Інженер.
5. Експерт (Експерт - фахівець, що дає висновок при розгляді питання.).
6. Робоча (оперативна) пам'ять.
7. База знань.
8. Вирішувач (механізм висновку).
9. Підсистема пояснень.

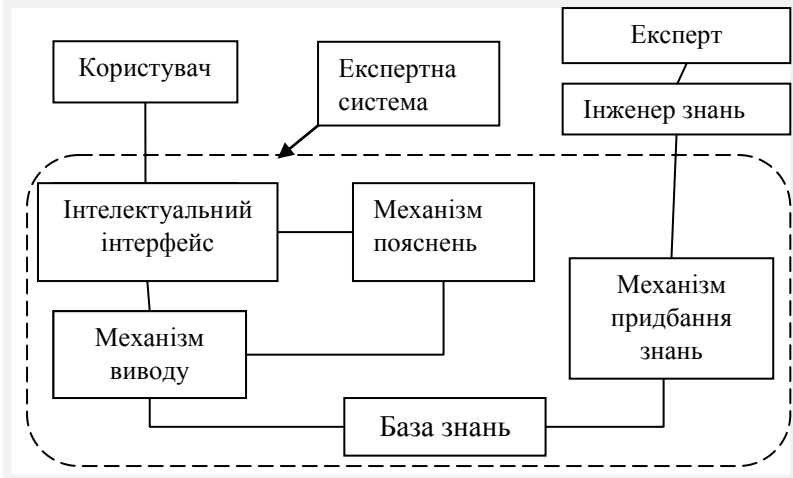


Рисунок 1 – Структура експертної системи

Етап структуривання знань - вибираються ІС і визначаються способи подання всіх видів знань, формалізуються основні поняття, визначаються способи інтерпретації знань, моделюється робота системи, оцінюється адекватність цілям системи зафіксованих понять, методів рішень, засобів представлення й маніпулювання знаннями.

Етап формалізації - здійснюється наповнення експертом бази знань. У зв'язку з тим, що основою ЕС є знання, даний етап є найбільш важливим і найбільш трудомістким етапом розробки ЕС. Процес придбання знань поділяють на вилучення знань з експерта, організацію знань, що забезпечує ефективну роботу системи, і представлення знань у вигляді, зрозумілому ЕС. Процес придбання знань здійснюється інженером по знаннях на основі аналізу діяльності експерта з вирішення реальних завдань.

Реалізація ЕС - створюється один або декількох прототипів ЕС, вирішальні необхідні завдання.

Етап тестування – проводиться оцінка обраного способу представлення знань в ЕС в цілому.

Для розробки експертних систем використовуються і звичайні мови програмування загального призначення: С, С++, С#, Java та ін.

Загальним недоліком мов програмування для створення експертних систем є: великий час розробки закінченої системи, необхідність залучення висококваліфікованих програмістів, труднощі з модифікацією закінченої системи. Все це робить застосування мов програмування для реалізації ЕС досить дорогим і трудомістким.

Спираючись на вищевикладене можемо зробити висновок, що експертні системи мають такі основні характеристики:

- експертна система обмежена певною сферою експертизи;
- вона здатна міркувати при сумнівних даних;
- вона здатна пояснити ланцюжок міркувань зрозумілим способом;
- вона будується так, щоб була можливість поступового нарощування системи;
- найчастіше вона заснована на використанні правил.

Отже, запропонована експертні системи управління мають свої переваги та недоліки, проте саме завдяки даним системам вирішуються безліч формалізованих і не формалізованих задач. Для прикладу використання експертних систем може бути застосовано для керування навчальним закладом та його процесом у реальному часі.

Подальші дослідження спрямовані на розробку методики впровадження експертних систем управління освітніми закладами.

Література

1. Красилів А.А. Основи інформатики. Визначення та концепції //Навчальний посібник, МФТІ, - М., 1990. - 80 с.
2. Джозеф Джарратано, Гари Райли, Експертные системы. Принципы разработки и программирование // 4-е издание / Вильямс, 2007. - 1152 с.
3. Попов Е.В. Експертні системи. Рішення неформалізованих завдань у діалозі з ЕОМ. - М.: Наука, 1987. - 288 с.
4. Уотермен Д. Керівництво по експертним системам: Пер. з англ. - М.: Світ, 1989. - 388 с.

Адаптивний метод аналізу кольорових зображень при нейромережевому розпізнаванні зашумлених образів

Ніколайчук В.А., Мазурець О.В.

Хмельницький національний університет

На сучасному етапі є актуальним завдання оперативного введення графічної інформації, підлягаючої подальшій обробці, в інформаційні і управляючі комп'ютерні системи. Проблема ефективного розпізнавання образів має важливе значення і в сферах автоматизації ряду процесів людської діяльності, пов'язаних з ідентифікацією різних об'єктів

навколишнього світу, наприклад, авторизація робочого персоналу по відбитках пальців або сітківці ока, ідентифікація продукту і розрахунок ціни в магазині по штрих-коду й ін..

Особливу складність і перспективність має завдання аналізу спектральної інформації для оптимізації розпізнавання зашумлених образів. Значну цінність для розв'язку даної проблеми додає можливість розпізнавання образів із неоднорідним фоном та автоматичне очищення образу від частини шумів шляхом його спектрального аналізу [1].

Метою роботи є розробка ефективного методу аналізу спектральної інформації з метою автоматизованого адаптивного розділення вхідної інформації на образ, що розпізнається, та фонове зображення. Областю застосування розроблених методів аналізу спектральної інформації визначено розпізнавання образів із неоднорідним фоном, значним ступенем спектрального спотворення, автоматичного очищення образу від частини шумів шляхом його спектрального аналізу.

Оскільки зображення містять великий об'єм інформації, важливу роль починають грати питання їх представлення. На практиці образи, що розпізнаються, мають спотворений вигляд. Є декілька основних видів спотворення, що суттєво впливають на можливість їх розпізнавання:

- площа і положення образу;
- орієнтація образу на зображенні;
- розмиття;
- випадання частин образу;
- зашумленість образу;
- зсув тощо.

Площа і положення образу має важливе значення для систем розпізнавання, так як в залежності від того, як око бачить об'єкт, залежить можливість його правильної ідентифікації.

Часто на практиці, важливим питанням є визначення орієнтації образу на зображенні. Об'єкт може бути повернутим в будь яку сторону, що суттєво вплине на результат роботи систем з використання штучного інтелекту.

При розпізнаванні, наприклад, відсканованих зображень, або старих документів досить часто спостерігається випадання частин образу, велика зашумленість, розмиття та зсув його частин. Тому аналіз і шляхи вирішення даних проблем мають велике практичне значення для успішного створення систем розпізнавання різного роду об'єктів навколишнього середовища.

Способи опису і обробки кольору відрізняються один від одного тим, для якого кінцевого уявлення призначаються. Порівняємо, наприклад, представлення кольорів для поліграфії і для моніторів комп'ютерів. У першому випадку за основу береться білий колір паперу, на який надалі наносяться три основні кольори: блакитний, пурпурний і жовтий. Змішуючись між собою і з білим кольором паперу в різних пропорціях, ці три

основні кольори дають різні колірні відтінки, окрім чистого чорного, або при повній відсутності фарб дають білий колір паперу. Якщо до них додати ще і чорний колір, то отримаємо СМУК-спосіб передачі кольору, коли необхідний колір виходить шляхом віднімання з білого бракуючих кольорів.

У другому ж випадку за основу береться чорний колір екрану монітора, кожен осередок якого, світиться одним з трьох кольорів: red-червоний, green-зелений і blue-синій. Тоді при повній відсутності якого-небудь свічення ми отримуємо чистий чорний колір екрану, а будь-який з необхідних кольорів задається співвідношенням кожного з трьох кольорів. В цьому випадку ми отримаємо RGB-спосіб передачі кольору. Основні кольори можуть мати значення від 0 до 255, або від 0% до 100%, або можуть бути представлені у вигляді шістнадцяткового значення.

Зображення в даній колірній моделі складається з трьох каналів. При змішенні основних кольорів (основними кольорами вважаються червоний, зелений і синій) — наприклад, синього (B) і червоного (R), ми отримуємо пурпурний (M magenta), при змішенні зеленого (G) і червоного (R) — жовтий (Y yellow), при змішенні зеленого (G) і синього (B) — ціаністий (Z cyan). При змішенні всіх трьох колірних компонентів ми отримуємо білий колір (W).

Аналіз спектральної інформації можна проводити по наступним основним критеріям:

- растрове спотворення фону;
- градієнтне спотворення фону;
- растрове спотворення образу;
- градієнтне спотворення образу;
- чистий образ;
- чистий фон.

При змішуванні усіх видів спектральних критеріїв з'являється велика кількість можливих комбінацій кольорів на зображенні.

Особливу складність і перспективність має завдання аналізу спектральної інформації для оптимізації розпізнавання зашумлених образів. Значну цінність розв'язку даної проблеми додає можливість розпізнавання образів із неоднорідним фоном та автоматичне очищення образу від частини шумів шляхом його спектрального аналізу. Метою роботи є розробка ефективного методу аналізу спектральної інформації з метою автоматизованого адаптивного розділення вхідної інформації на образ, що розпізнається, та фонове зображення.

Тому, зважаючи на специфіку області застосування (необхідність автоматизованого адаптивного розділення вхідної інформації, варіантність можливих розв'язків), переваги системного підходу й перспективність методу динамічної корекції і самовдосконалення виконавчого механізму, в якості технології штучного інтелекту для розробки технології аналізу спектральної інформації було обрано апарат нейронних схем.

Обов'язковою вимогою роботи з кольоровими патернами образів є мультифункціональність і можливість варіантного прийняття рішень. Тому в рамках розробки нейросхемного методу [2] аналізу спектральної інформації реалізовано два способи розподілу спектральної інформації (Рисунок 1).

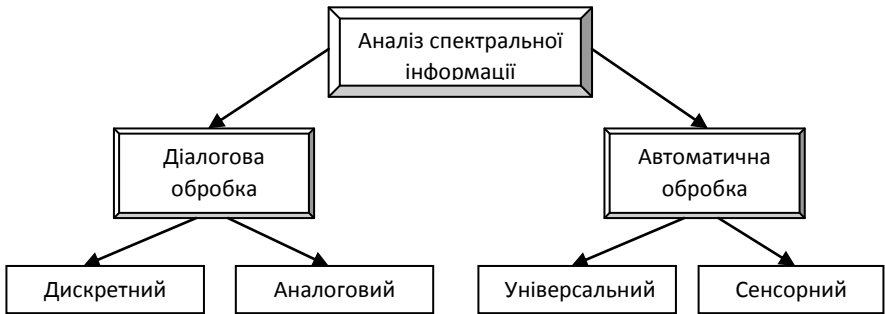


Рисунок 1 – Загальна структура варіантів аналізу спектральної інформації

Діалогове визначення спектрального балансу – це ручні методи аналізу та обробки спектральної інформації. Можливі два способи реалізації цього методу:

- дискретний – вибір кольорових патернів;
- аналоговий – вибір спектральної межі.

Автоматизований аналіз спектрального розподілу реалізується шляхом ітераційного аналізу поступаючої від нейросхеми інформації в процесі розподілення кольорового балансу й включає наступні варіанти виконання:

- однофазний – тестування образу нейросхемою на знаходження локальних максимумів вірогідності розпізнавання;
- самокорегувальний – автоматичне прийняття результату роботи однофазного етапу як еталонної моделі.

Для реалізації дискретного методу аналізу спектральної інформації було використано людську можливість визначення кольорів. В цілому, зображення може складатися з багатьох різних кольорів, але всі вони можуть бути поділені на два типи: ті, що належать фону і ті, що зображують образ. Тому, розділивши на палітрі, кольорові патерни на фон та образ, можна подавати оброблене зображення на розпізнавання нейросхемою.

Недоліком даного методу є те, що не завжди є можливість правильно розділити всі кольори зображення на фон та образ, що призведе до втрати ефективності при роботі нейронної схеми. При значно зашумленому зображенні використання дискретного методу взагалі втрачає свою необхідність.

Аналоговий метод аналізу спектральної інформації реалізований на основі попереднього. Залежність від користувача при визначенні спектру кольорів, робить його не досить ефективним. Для його реалізації було розроблено дві спектральні межі, одна відповідає за кольори, що належать

фону, інша – образу. У випадку, коли зображення, що подається на розпізнавання має невелику ступінь зашумленості, даний метод є досить ефективним і дієвим. Аналоговий метод передбачає в собі певну автоматизу, оскільки не потрібно вказувати кожен колір окремо, можна, лише вказати спектр, що належить фону або образу. Ефективність роботи аналогового методу залежить від ступеня спотвореності вхідного зображення. При значній спектральній зашумленості і розкиду гами кольорів на зображенні, використання методу стає досить незручним, хоча при розпізнаванні зображень із чистим фоном або чистим образом дає досить непогані результати і може бути застосованим на практиці. Даний спосіб обробки вхідної спектральної інформації часто використовується в Photoshop та інших прикладних програмах, тому його розробка є одним із ланок, що з'єднує діалогові та автоматизовані методи.

Для автоматичного розпізнавання образів реалізовано також два способи. Відмінність їх полягає у швидкості роботи. Першим і найбільш затратним по часу є універсальний метод аналізу спектральної інформації.

Універсальний метод аналізу спектральної інформації передбачає повний перебір — метод рішення задачі шляхом перебору всіх можливих варіантів. Складність повного перебору залежить від розмірності простору всіх можливих рішень задачі. Якщо простір рішень дуже великий, то повний перебір може не дати результатів протягом декількох років або навіть століть. Будь-яке завдання з класу NP може бути вирішене повним перебором. При цьому, навіть якщо обчислення цільової функції від кожного конкретного можливого рішення задачі може бути здійснена за поліноміальний час, залежно від кількості всіх можливих рішень повний перебір може зажадати експоненціального часу роботи. У випадку роботи з зашумленими образами, елементами для перебору є фонові кольори. Використання нейронних схем, дало можливість не брати до уваги кольори, що належать образу. Використання універсального методу дає не погані результати, але з надзвичайно великими затратами часу, тому наступний метод частково вирішує цю проблему.

Сенсорний метод аналізу спектральної інформації також передбачає перебір всіх можливих варіантів. Але кількість кольорів, що потрібно перебрати, значно менша аніж при роботі універсального методу. На початковому етапі своєї роботи, визначаються всі кольори, що потрапляють в сітківку «ока», після чого, шляхом повного їх перебору визначається максимальний відсоток розпізнавання образу. Це є основним методом аналізу спектральної інформації, оскільки він значно швидший за універсальний метод, а по ефективності розпізнавання не поступається йому. Із збільшенням кількості кольорів шумів зменшується швидкість роботи методу. На практиці, в межі сітківки «ока» попадає 3-5 різних кольорів, що суттєво впливає на роботу методу. Сенсорний метод аналізу спектральної

інформації зашумлених зображень в поєднанні з нейросхемними алгоритмами розпізнавання спотворених образів дає найкращі результати серед наведених вище, тому його аналіз і подальша доробка є актуальним питанням. Сенсорний метод по суті виконує ті ж функції, що і універсальний, але швидкість його роботи залежить від кількості кольорів фону.

Таким чином, було запропоновано новий метод аналізу спектральної інформації з метою автоматизованого адаптивного розділення вхідної інформації на образ, що розпізнається, та фонове зображення. Розроблений метод має значну цінність для розв'язку проблеми розпізнавання образів. Для перевірки ефективності розробленого методу створена система, що реалізує розроблений метод спектрального аналізу образів у порівнянні з існуючими методами, яка підтвердила його ефективність.

Призначенням методу аналізу спектральної інформації є: розпізнавання образів із неоднорідним фоном; розпізнавання образів із значним ступенем спектрального спотворення; автоматичне очищення образу від частини шумів шляхом його спектрального аналізу.

Література

1. Ковальчук С.С., Мазурець О.В. Розробка технології декомпозитивного розпізнавання символічної інформації з креслень на базі нейронних схем // Прикладна геометрія та інженерна графіка. Праці Таврійського державного агротехнологічного університету / Таврійський державний агротехнологічний університет. Вип.4, т.44. – Мелітополь: ТДАТУ, 2009. – С.86–94.

2. Мазурець О.В. Розробка автоматизованої системи для розпізнавання растрових креслень на базі нейронних схем // Збірник наукових праць за матеріалами другої всеукраїнської науково-технічної конференції «Актуальні проблеми комп'ютерних технологій 2008» – Хмельницький – ХНУ, 2008. – Том 2 – С.22–27.

Інформаційна система забезпечення ефективності функціонування бізнес-процесів організації на прикладі кінотеатру

Пшедзьял Я.Е, Скрипник Т.К.

Хмельницький національний університет

У сучасному світі вплив інформаційних технологій на всі аспекти життя є дуже високим, у тому числі на сфері послуг, розваг, підприємництва, освіти. Доцільність впровадження інформаційних технологій у кінотеатрах обґрунтовується зростаючою потребою у високому рівні сервісу та обслуговування відвідувачів, потребі в оптимізації бізнес-процесів організації та впровадженні новітніх маркетингових прийомів та підприємницьких підходів [1].

Базовою складовою існуючих інформаційних автоматизованих систем будь-якого кінотеатру є набір функцій для обліку продажу квитків працівниками. Ця частина залишається однією з найбільш важливих та технічно складних, незважаючи на трансформацію потреб до таких систем. При цьому стрімка модернізація, виникнення потреби у можливості самостійно придбати квиток на публічній платформі, нові сервіси, а також потреба в аналізі й опрацюванні даних кінотеатру породжують нові завдання проектування, такі як: забезпечення можливості масштабування для багатьох джерел інформації (з інтернет-сайту, каси, персональних пристроїв), об'єднання мереж кінотеатрів у розподілені системи, оптимізація підтримки апаратно-програмної інфраструктури.

Потенційно, у найближчому майбутньому, поставатимуть нові й нові функціональні вимоги до таких автоматизованих систем. Наприклад, інтелектуальний аналіз «великих даних» кінотеатру може стати невід'ємною частиною бізнес-процесів кінотеатру, пов'язаних з плануванням, дослідженнями, маркетингом [2]. Існують загальні для будь-якої організації бізнес-процеси, що також потребують автоматизації: адміністрування, облік та звітність. У випадку кінотеатру автоматизація формування звітності для кінодистриб'юторів та державних регулюючих органів є особливо важливою, оскільки потребує зберігання та опрацювання величезного обсягу даних.

Метою роботи є розробка інформаційної системи кінотеатру, а також її ефективної програмно-апаратної платформи та архітектури, що буде здатна масштабуватись та змінюватись відповідно до нових та існуючих бізнес-процесів організації, забезпечуватиме стабільну роботу навіть при високих технічних навантаженнях й залишатиметься гнучкою для впровадження нових функцій.

Функціональні вимоги до автоматизованої системи кінотеатру можна поділити на кілька груп залежно від прогнозованої ресурсоемності операцій та частоти їх використання:

1. Ресурсоемні, періодичні операції аналізу, архівування даних, звітності.
2. Нересурсоемні, часто використовувані функції роботи з сайтом, його адміністрування.
3. Найбільш часто використовувані операції, пов'язані з транзакціями купівлі квитків та завантаженням, трансляцією інформаційних ресурсів.

Окремо до таких систем ставлять нефункціональні вимоги, як легкість підтримки серверів, програмного коду та програмно-апаратної інфраструктури, швидкість роботи системи, адаптивність до рівню навантажень, економічна доцільність, надійність збереження даних.

Обираючи програмну платформу та архітектуру майбутньої інформаційної системи кінотеатру, в першу чергу варто проаналізувати варіант класичної «монолітної» архітектури, що зображена на рисунку 1.

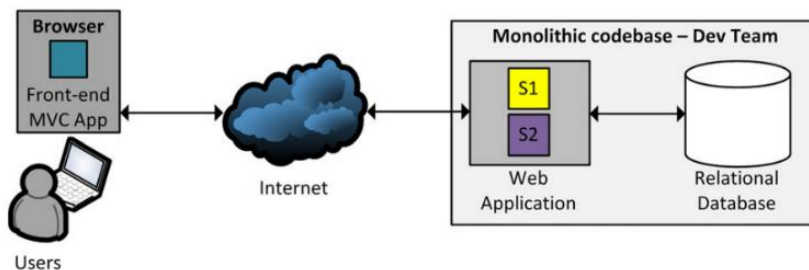


Рисунок 1 – «Монолітна» архітектура веб-додатку

Вона дозволяє будувати великі програмні додатки, які складаються з безлічі функціональних частин в рамках однієї великої структури. Існує багато прогнаних платформ, що розроблені саме для такої архітектури: наприклад ASP .NET MVC, Symfony, Rails. Усі вони мають перевірені часом інструменти та рішення для подібних задач, інтеграції з сторонніми сервісами, як-от платіжними системами, а також автоматизації складних бізнес-процесів, при цьому залишаючись гнучкими та легкими у підтримці. Перевагами цього підходу є незалежність від будь-яких сторонніх ресурсів, з технічної точки зору – його структурованість, надійність та передбачуваність у роботі. Проте такий підхід часто не здатний забезпечити потрібну ефективність для завдань та процесів, яким характерні непередбачуваність, нетипово високі або навпаки, низькі навантаження або таких, де потоки даних та взаємодії з часом змінюються. Це пов'язано з тим, що монолітна архітектура сильно пов'язана з апаратною частиною. Як правило, її робота забезпечується єдиним централізованим сервером, а в разі потреби розширення та масштабування, несе з собою витрати на модернізацію та підтримку апаратної інфраструктури [3].

На противагу йому можливо застосувати сучасну мікросервісну або “serverless” архітектуру, що базується на «хмарних» технологіях. Такий підхід суттєво відрізняється від попереднього і полягає він у використанні багатьох окремих, дуже маленьких програмних компонент, кожен з яких є окремим веб-додатком [4]. Кожному компоненту даної архітектури відповідає певна роль у функціонуванні програми, а підходу в цілому характерні автономність, слабкі пов'язаність та інтегрованість (рис. 2).

В якості апаратної інфраструктури таких систем виступають віддалені сервери, доступ до яких забезпечує сама технологія на умовах платного сервісу. Як наслідок, перевагою підходу є гнучкість системи, можливість вдаватися до автоматичного, миттєвого масштабування окремих компонент (рис. 3). Такі додатки не потребують підтримки й адміністрування апаратної частини – серверів, є надзвичайно масштабованими й економічно доцільними через ще одну особливість: плата відбувається виключно за використані ресурси.

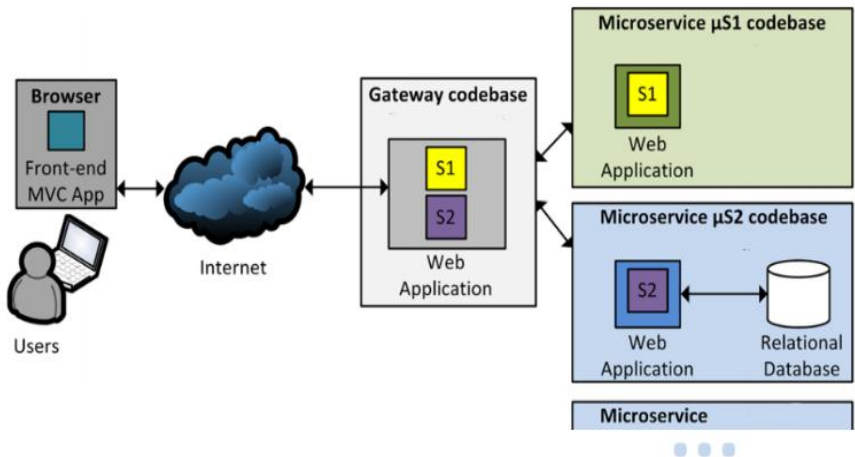


Рисунок 2 – Мікросервісна архітектура веб-додатку

Застосування такої архітектури накладає деякі обмеження на технологічні рішення, що застосовуватимуться, позбавляє інформаційну систему цілісності, структурованості, ускладнює процес розробки та тестування. Також для «хмарної» архітектури характерний негативний ефект «холодного старту», проілюстрований на рисунку 4. Він полягає у тому, що ті частини системи, що використовуються не стабільно часто, можуть потребувати більше часу, ніж зазвичай, для первинної ініціалізації. Інколи це може становити кілька секунд, що створює незручності для користувачів інформаційної системи.

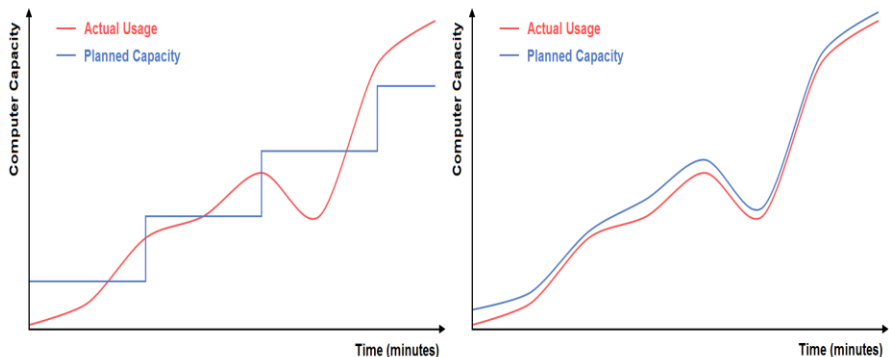


Рисунок 3 – Планування обчислювальної потужності за використання монолітних платформ (зліва) та застосовуючи Serverless «хмарні» технології (справа)

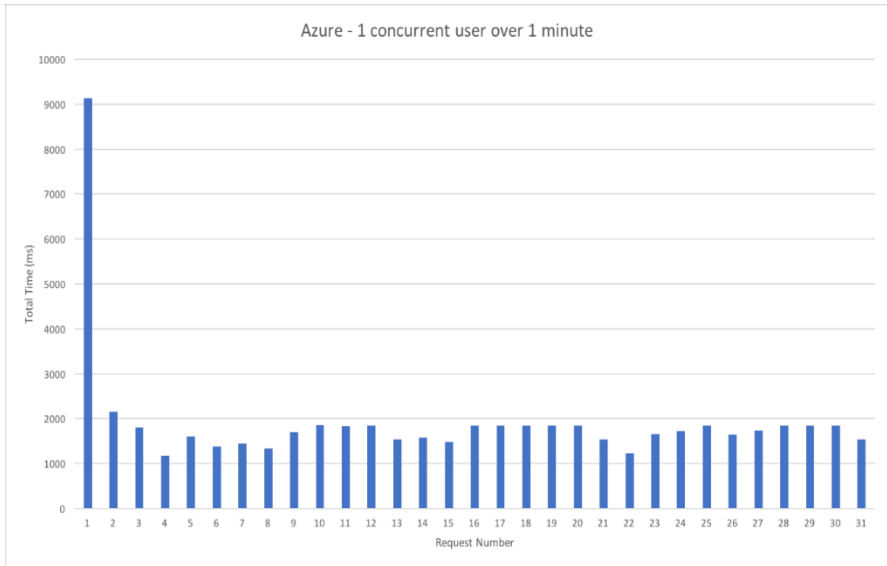


Рисунок 4 – Ефект «холодного старту» мікросервісу

Підсумувавши особливості кожної з платформ та архітектур, з огляду на специфіку бізнес-процесів предметної області кінотеатру, вибір лише одного з розглянутих підходів є недостатнім для забезпечення найбільш ефективного функціонування бізнес-процесів організації. Функції розроблюваної інформаційної системи є досить різносторонніми, тому для отримання найкращого архітектурного рішення доцільно створити нову інформаційну технологію, що комбінуватиме обидва розглянутих підходи (рис. 5) й розподілятиме інформаційні потоки у системі між ними для кожного специфічного бізнес-процесу, що можна спостерігати на рисунку 6.

Таким чином, пропонується функції інформаційної системи, що забезпечують бізнес-процеси функціонування каси, онлайн продажу квитків автоматизувати, розробивши окремі мікросервіси, виділивши для їх функціонування найбільший пріоритет та вільне масштабування обчислювальних потужностей. Це дозволить системі залишатись швидкою та стабільною незалежно від навантаження, оскільки ці функції інформаційної системи є найбільш часто використовуваними.

Цей підхід доречно застосувати і для організації доступу користувачів до інформаційних ресурсів та мультимедійного контенту, наприклад опису кінофільмів та їх відео трейлерів, що будуть доступні в мережі Інтернет.

Саме для таких потреб Serverless платформа надає зручний інструментарій – можливість створити CDN (Content Delivery Network), що розповсюджуватиме мультимедійні файли.

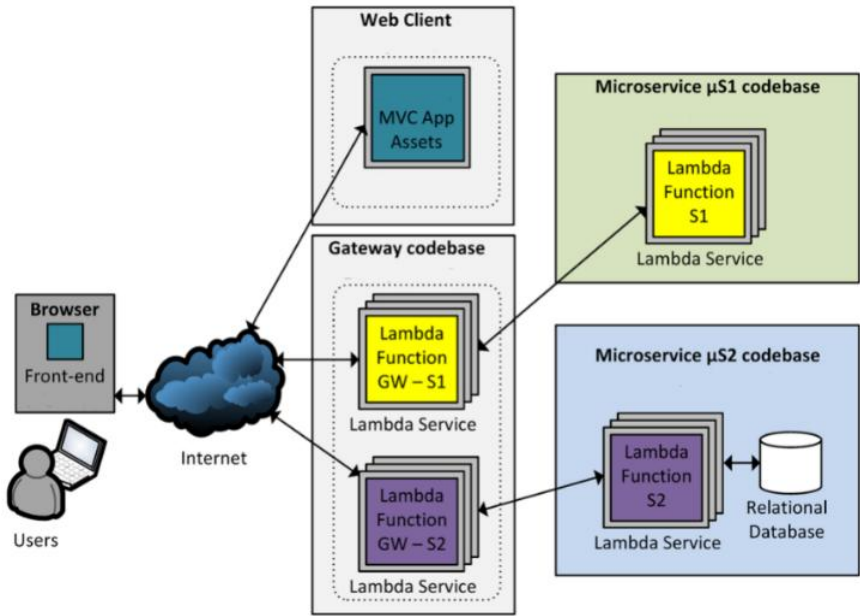


Рисунок 5 – Комбінована архітектура, що поєднує мікросервіси та «монолітний» MVC додаток в якості серверної частини.

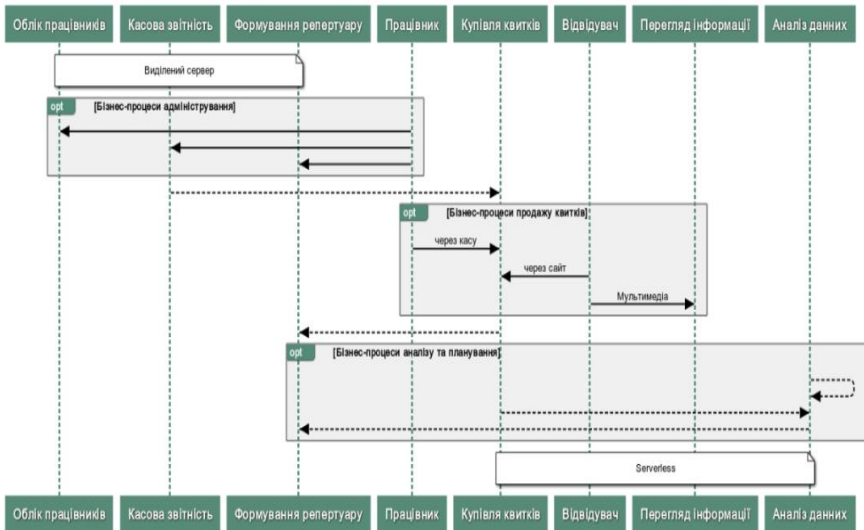


Рисунок 6 – Розподіл бізнес-процесів між інформаційними підсистемами

Для автоматизації бізнес-процесів пов'язаних з адмініструванням інформаційної системи та самого кінотеатру організуємо окремий програмний додаток за класичною, монолітною архітектурою. Така підсистема забезпечить надійність, передбачуваність та цілісність процесів та даних у ній, при цьому залишатиметься простою та гнучкою, а оскільки вона використовуватиметься переважно персоналом кінотеатру або кіномережі, проблема високих або непередбачуваних навантажень для неї не є актуальною.

Отже, щоб забезпечити високу ефективність функціонування бізнес-процесів організації, а саме кінотеатру, за допомогою інформаційних технологій, були проаналізовані існуючі підходи, програмні та архітектурні платформи для розробки інформаційних систем і як результат, було запропоновано нова інформаційна технологія, що комбінує «монолітну» клієнт-серверну архітектуру та «хмарну» serverless платформу для автоматизації специфічних бізнес-процесів та вимог, що мають місце у даній предметній області.

Література

1. Mindi Chahal How innovation and technology are changing the cinema experience for brands [Електронний ресурс]. – Режим доступу: <https://www.marketingweek.com/2016/09/28/how-innovation-and-technology-are-changing-the-cinema-experience-for-brands/>

2. Jackie Y. Luan Optimal Inter-Release Time between Sequentially Released Products / Jackie Y. Luan, K. Sudhir // Yale University, 2016 – 38 с.

3. Якимець Р. В. Масштабування навантаження Web-додатків / Р. В. Якимець, К. М. Яременко // Міжнародний науковий журнал, 2016, № 6(2). - С. 45-47.

4. Г.Ю. Сидоренко Використання Serverless підходу для створення веб-додатку моніторингу товарів / Г.Ю. Сидоренко, М.М. Малько, М.А. Ляшенко // Вісник Національного технічного університету «ХПІ». Серія: Системний аналіз, управління та інформаційні технології, № 21 (1297), 2018

Інформаційна система автоматизованого реферування

П'ятін В.Д., Свірневський М.С.

Хмельницький національний університет

Автоматичне реферування (Automatic Text Summarization) - це складання коротких викладів матеріалів, анотацій або дайджестів, тобто витяг найбільш важливих відомостей з одного або декількох документів і генерація на їхній основі лаконічних звітів.

Існує багато шляхів розв'язку цього завдання, які досить чітко підрозділяються на два напрямки - квазіреферування та короткий виклад змісту первинних документів [1]. Квазіреферування засноване на екстрагуванні фрагментів документів – виділенні найбільш інформативних

фраз і формуванні з них квазірефератів.

У рамках квазіреферування виділяють три основні напрямки, які в сучасних системах застосовуються спільно:

- статистичні методи, засновані на оцінці інформативності різних елементів тексту по частоті появи, яка служить основним критерієм інформативності слів, пропозицій або фраз;

- позиційні методи, які опираються на припущення про те, що інформативність елемента тексту залежить від його позиції в документі;

- індикаторні методи, засновані на оцінці елементів тексту, виходячи з наявності в них спеціальних слів і словосполучень - маркерів важливості, які характеризують їхню змістовну значимість.

Метою роботи є проведення оцінки ефективності автоматизованого реферування на базі існуючих підходів та застосування багаторівневих підходів.

Визначення ваги фрагментів (пропозицій або абзаців) вихідного тексту виконується відповідно до алгоритмів, які стали вже традиційними.

Location визначається розташуванням блоку в тексті та залежить від того, де з'являється даний фрагмент - на початку, у середині або наприкінці, а також чи використовується він у найбільш важливій зі змістовної точки зору розділах тексту, наприклад, у висновках [2]. Ключові фрази (Keuphrase) являють собою конструкції-маркери, які резюмують зміст, типу "на закінчення", "у даній статті", "у результаті аналізу" і т.п. Вагове значення, що складається Keuphrase може залежати також від оцінного терміна, наприклад, "відмінний". Статистична вага текстового блоку (Statterm) обчислюється як нормована по довжині блоку сума ваг вхідних у нього слів і словосполучень.

Після виявлення певної (що задається, як правило, коефіцієнтом необхідного стиску) кількості текстових блоків з найвищими ваговими коефіцієнтами, вони поєднуються для побудови квазіреферата.

Перевага методів квазіреферування полягає в простоті їх реалізації. Однак виділення текстових блоків, що не враховує взаємин між ними, часто приводить до формування нескладних рефератів. Деякі теми можуть виявитися пропущені, або в них можуть зустрічатися слова або фрази, які неможливо зрозуміти без попереднього пропущеного тексту. Спроби розв'язати цю проблему, в основному зводяться до виключення таких пропозицій з рефератів. Рідше робляться спроби посилянь за допомогою методів лінгвістичного аналізу.

Короткий виклад змісту первинних документів ґрунтується на виділенні з текстів найбільш важливої інформації та породженні нових текстів первинних документів, що змістовно узагальнюють. На відміну від частотно-лінгвістичних методів, що забезпечують квазіреферування, підхід, заснований на базах знань, опирається на автоматизований якісний контент-аналіз, що полягає, як правило, із трьох основних стадій. Перша - відомість вихідної текстової інформації до заданого числа фрагментів - одиниць

значення, якими є категорії, послідовності та теми. На другій стадії проводиться пошук регулярних зв'язків між одиницями значення, після чого починається третя стадія - формування висновків і узагальнень. На цій стадії створюється структурна анотація, що представляє зміст тексту у вигляді сукупності концептуально зв'язаних значеннєвих одиниць.

Семантичні методи формування рефератів-викладів припускають два основні підходи: метод синтаксичного розбору пропозицій і методи, що опираються на розуміння природньої мови. У першому випадку використовуються дерева розбору тексту. Процедури автоматичного реферування маніпулюють безпосередньо деревами, виконуючи перегрупування та скорочення галузей на підставі відповідних критеріїв. Таке спрощення забезпечує побудова реферату - структурну "вичавлювання" вихідного тексту.

Другий підхід ґрунтується на системах штучного інтелекту, у яких також на етапі аналізу виконується синтаксичний розбір тексту, але синтаксичні дерева не породжуються. У цьому випадку формуються семантичні структури, які накопичуються у вигляді концептуальних підграфів у базі знань. Зокрема, відомі моделі, що дозволяють робити реферування текстів на основі психологічних асоціацій подібності та контрасту. У базах знань надлишкова, що не має прямого відношення до тексту інформація усувається шляхом відсікання деяких підграфів. Потім інформація зазнає агрегування методом злиття, що залишилися графів або їх узагальнення. Для виконання цих перетворень виконуються маніпуляції логічними припущеннями, виділяються визначальні шаблони в текстовій базі знань. У результаті перетворення формується концептуальна структура тексту - анотація, тобто концептуальні "вичавки" з тексту.

Багаторівневе структурування тексту з використанням семантичних методів дозволяє застосовувати підходити та вирішувати завдання реферування таким шляхом:

- видалення малозначних одиниць. Перевагою методу є гарантоване збереження значущої інформації, недоліком - низький ступінь стиску, тобто скорочення обсягу реферату в порівнянні з первинними документами;
- скорочення значимих одиниць - заміна їх основною лексичною одиницею, що виражає основний зміст;
- гібридного способу, що полягає в уточненні реферату за допомогою статистичних методів, з використанням семантичних класів, особливостей контексту та синонімічних зв'язків.

Існують загальнодоступні програми квазіреферування, наприклад, в складі сервісних можливостей системи Microsoft Word є режим "Автореферат".

Інформаційна система використовує підхід без вчителя до реферування тексту на основі графів підрахунку центрування змісту. Головна ідея полягає в тому, що пропозиції "рекомендують" інші подібні речення

читачеві. Отже, якщо одне речення дуже подібне до багатьох інших, воно, швидше за все, стане предметом визначального значення. Важливість цього речення також впливає з важливості висновків, які "рекомендують" це. Таким чином, щоб отримати високий рейтинг і помістити у резюме, речення має бути схожим на речення, які, в свою чергу, також схожі на багато інших речень. Це дозволяє знайти інтуїтивний сенс і застосовувати алгоритми до будь-якого довільного нового тексту.

Література

1. Luhn H.P. The automatic creation of literature abstracts // IBM J. of Research and Development. 1958. 2. N 2. P.159-165.
2. Яцко В.А. Симметричное реферирование: теоретические основы и методика // НТИ. 2002. 2. N 5. С. 18-28.
3. Кліменко В.І., Живілік А.В., Мазурець О.В. Аналіз сучасних методів автоматизації анутовання та реферування текстів // Збірник наукових праць за матеріалами дев'ятої міжнародної науково-технічної конференції «Актуальні проблеми комп'ютерних технологій 2015». Хмельницький – 2015. – С.116-123.
4. Бармак О. В., Мазурець О. В., Живілік А. В. Інформаційна технологія автоматизованого анутовання та реферування цифрових текстів // Науковий журнал „Вісник Хмельницького національного університету” серія: Технічні науки. Хмельницький, 2017, №4. – С.147-158.
5. Kiss T., Strunk J. Unsupervised multilingual sentence boundary detection // Computational Linguistics. 2006. 32. N 4. P. 485–525.
6. Dunning T. Accurate methods for the statistics of surprise and coincidence // Computational Linguistics. 1993. 19. N 1. P. 61–74.
7. Müller H., Amerl V., Natalis G. Worterkennungs verfahren als Grundlage einer Universal methode zur automatischen Segmentierung von Texten in Sätze. Ein. 1980. 4. N 1. P. 46–64.

Мережеві технології забезпечення відображення віртуального ігрового контенту

Рибачук В.В.

Хмельницький національний університет

З розвитком мережевих технологій виникає потреба підвищення швидкості передачі інформації. Особливо важливою така проблема постає в індустрії комп'ютерних ігор, де швидкість передачі віртуального ігрового контенту стає лімітуючим фактором розвитку.

В мережевих технологіях відображення ігрового контенту використовують стандартизовані XML серіалізатори і десеріалізатори для передавання даних. В кожній із технологій не вистачає має достатнього функціоналу для повноцінної роботи, яка потрібна для реалізації задач які постають перед розробниками ігрового контенту.

Мережеві технології складаються з декількох транспортних рівнів, в роботі використовують 4 та 5 рівень моделі OSI, призначений для доставки даних без помилок, втрат і дублювання в тій послідовності, якій вони були передані. При цьому неважливо, які дані передаються, звідки й куди, тобто надає сам механізм передачі. Блоки даних розділяють на фрагменти, розмір яких залежить від протоколу, короткі поєднують в один, великі розбивають. Протокол транспортного рівня може не використовувати усі вищевказані можливості.

При передачі даних важливою частиною є серіалізація та десеріалізація об'єктів. Серіалізація використовується для шифрування даних контенту. Для відновлення пересланих даних зазвичай використовують десеріалізацію, яка відбувається на стороні хоста-отримувача або хоста-відправника.

Для шифрування даних використовуються декілька основних методів:

- метод здійснення віддалених викликів процедур;
- метод розповсюдження об'єктів, особливо в технологіях

компонентно-орієнтованого програмування;

- метод виявлення змін у даних, що змінюються з часом.

Аналіз мережевих технологій відображення ігрового контенту показав, що алгоритми передачі даних мають ряд недоліків. Недоліки зв'язані з швидкістю передачі даних, що в реаліях ігрового контенту дуже важливо, так як швидкість впливає на затримки під час ігрової сесії.

З огляду на вище вказане, основною метою роботи збільшення швидкості передачі даних ігрового контенту, шляхом вдосконалення алгоритмів шифрування даних та модифікації технологій. Основним напрямком для реалізації вище вказаних пропозицій було обрано бінарну серіалізацію та десеріалізацію даних. Використання бінарної серіалізації та десеріалізації даних підвищує швидкість шифрування та передачі інформації, що зменшує час відгуку від сервера до хоста-приймача.

Для аналізу ефективності запропонованих алгоритмів підвищення швидкості передачі ігрового контенту в мережі, було реалізовано програмний продукт в якому реалізовано бінарну технологію серіалізації та десеріалізації даних.

Для підтвердження запропонованого підходу для підвищення швидкості передачі ігрового контенту було проведено дослідження з використанням технологій UNET та запропонованою бінарною технологією. Вибір обумовлено тим, що в UNET використовується XML серіалізація та десеріалізація передачі даних в мережі відображення ігрового контенту.

Результати порівняльного аналізу базової та запропонованої технологій наведено в таблиці. 1. Аналіз показує, що для всіх типів даних відбувається збільшення швидкості передачі даних за рахунок збільшення швидкості серіалізації та десеріалізації даних.

Таблиця 1 – Порівняння технологій

	Технологія UNET	Бінарна технологія
Array	Avg 1.66337 ms.	Avg 0.66337 ms.
String	Avg 0.11337 ms.	Avg 0.056337 ms.
Int, float, long, byte	Avg 0.1801 ms.	Avg 0.0901 ms.

Таким чином, завдяки запропонованим алгоритмам, які реалізовані в бінарній технології серіалізації та десеріалізації даних, отримано суттєве зменшення часу передачі в мережі ігрового контенту в порівнянні з технологією UNET.

Література

1. Melville P., Mooney R., Nagarajan R. Content-Boosted Collaborative Filtering for Improved Recommendations// University of Texas, USA : Матеріали конф. / AAAI-02, Austin, TX, USA, 2002. — 2002. — P. 187-192.
2. Matrix Factorization Techniques for Recommender Systems. Computer (IEEE) 42 (8): 30–37. – Режим доступу: Koren, Y.; Bell, R.; Volinsky, C. (07 August 2009).
3. Lika, Blerina; Kolomvatsos, Kostas; Hadjiefthymiades, Stathes (March 2014). "Facing the cold start problem in recommender systems". Expert Systems with Applications.

Використання методу ансамблів для класифікації

Рижак А.В., Лищук О.А.

Хмельницький національний університет

До класу комітетних методів класифікації відносяться методи, у яких при ухваленні рішення про те до якого класу відноситься представлений опис об'єкта, бере участь кілька класифікаторів.

Є багато аналогій між комітетними методами та колективним ухваленням рішення, тому часто розв'язок про клас ухвалюється за результатами «голосування» класифікаторів, коли підраховується число класифікаторів, що проголосували за кожний клас, і визначається більшість [1-3]. Ще одним розповсюдженим варіантом є усереднення вихідних сигналів класифікаторів і ухвалення рішення на основі отриманого результату. Природно, є більш складні схеми комітетної класифікації, наприклад, зважена згортка розв'язків окремих класифікаторів, або багаторівневе голосування. Проте, оптимальної схеми комітетної класифікації немає.

Відзначимо, що нейронна мережа також може вважатися комітетним методом, якщо вважати, що кожний нейрон – це окремий класифікатор, який уміє розпізнавати два класи. Тоді вся мережа, це багаторівнева композиція таких бінарних класифікаторів. Однак, у даній роботі будемо розглядати як

класифікатор «цілком».

В одному комітеті можуть бути класифікатори різного типу, наприклад, нейронні мережі, дерева рішень та машини опорних векторів. Застосування комітетних методів теоретично не гірше застосування одного класифікатора. Це правило часто спостерігається на практиці, однак бувають випадки, коли комітетна класифікація працює гірше одного класифікатора. Тому корисність комітетної класифікації для розв'язку конкретного завдання, як правило, визначається експериментально.

Мета роботи полягає у застосуванні методу ансамблів при класифікації та оцінюванні його ефективності.

Опишемо методи комітетної класифікації: беггінг і бустинг.

Беггінг (Bagging) представляє один з базових алгоритмів, коли комітетний розв'язок ухвалюється усередненням розв'язків окремих класифікаторів.

Для навчання класифікаторів застосовується бутстреп (bootstrap) де навчальна вибірка розділюється на M рівних непересічних частин, по кількості використовуваних класифікаторів, і кожний класифікатор навчається без однієї частини даних. Тобто перший класифікатор може бути навчений без першої частини, другий – без другої і т.д. Таким чином, кожний класифікатор буде навчений на різних множинах даних.

В алгоритмі бустинга класифікатори навчаються послідовно, тобто навчання m -го класифікатора залежить від результату навчання попередніх $(m-1)$ класифікаторів. При цьому під час навчання m -го класифікатора більше уваги приділяється прикладам, на яких частіше помиляються попередні класифікатори. Після навчання комітету класифікація проводиться шляхом зваженого голосування, при цьому вага кожного класифікатора визначається залежно від результатів навчання. Даний підхід розроблявся для розв'язку завдань бінарної класифікації, із двома класами, але пізніше була запропонована модифікація алгоритму для довільного числа класів.

Далі буде розглянутий варіант для бінарної класифікації з довільним числом класифікаторів.

Для реалізації послідовного навчання кожному навчальному прикладу привласнюється вага, а в якості функції помилки застосовується функція виду:

$$err = \sum_i v_i [y_i \neq t_i], \quad (1)$$

де v_i – вага i -го навчального прикладу; y_i і t_i – відповідно фактичний і необхідний сигнал на виході; $[y_i \neq t_i]$ – індикаторна функція:

$$[y_i \neq t_i] = \begin{cases} 1, & y_i \neq t_i \\ 0, & y_i = t_i \end{cases} \quad (2)$$

Ваги прикладів оновлюються за результатами навчання кожного класифікатора, тобто при навчанні різних класифікаторів ваги прикладів у

загальному випадку будуть відрізнятися. Задамо величину

$$e_m = \frac{\sum_i v_i [y_i \neq t_i]}{\sum_i v_i}, \quad (3)$$

яка відображає нормалізовану помилку m -го класифікатора, що не залежить від масштабу ваг. Вага («якість») m -го класифікатора обчислюється як

$$\alpha_m = \ln \left(\frac{1 - e_m}{e_m} \right). \quad (4)$$

Таким чином, чим менше нормалізована помилка, тим більше вага класифікатора. На підставі величини ваги m -го класифікатора проводиться відновлення ваг прикладів для навчання $m+1$ класифікатора:

$$v_i^{(m+1)} = v_i^{(m)} \exp(\alpha_m [y_i \neq t_i]). \quad (5)$$

Для тих прикладів, які правильно розпізнаються m класифікатором, вага не змінюється. Для неправильно розпізнаних прикладів зміна ваги залежить від нормалізованої помилки класифікатора. При $e_m > 0.5$, тобто якщо якість класифікатора низька, вага $\alpha_m < 0$, і вага неправильно розпізнаного прикладу буде зменшуватися. При $e_m < 0.5$ вага $\alpha_m > 0$, і вага неправильно розпізнаного прикладу буде збільшуватися.

В алгоритмі бустинга як правило використовуються так звані «слабкі» алгоритми класифікації, для яких досить точності класифікації лише дещо краще випадкової класифікації. Враховуючи, що вище описаний варіант алгоритму для завдання бінарної класифікації, для слабого класифікатора досить точності розпізнавання більше 50%. Тобто e_m теоретично повинне бути менше 0.5, і в складних для розпізнавання прикладів вага повинна збільшуватися.

У цілому, алгоритм бустинга виглядає таким чином:

1. Ініціалізація ваг навчальних прикладів: $v_i = 1/N$, N – кількість навчальних прикладів.

2. Для кожного з M тих, яких навчають, класифікаторів:

- a. Навчання класифікатора з використанням цільової функції (1)
- b. Обчислення ваги класифікатора по формулі (2).
- c. Відновлення ваг навчальних прикладів (3).

3. Ухвалення рішення з використанням виразу:

$$y(x) = \sum_{m=1}^M \alpha_m y_m(x),$$

де $y_m(x)$ – вихідний сигнал m -го класифікатора для вхідного вектора X .

Застосування бустинга для навчання комітету нейронних мереж.

Оскільки (1) суттєво відрізняється від середньоквадратичної функції помилки, часто використовуваної при навчанні, то для застосування бустинга

для навчання комітету необхідно зробити одне із двох:

1. Внести зміни у формули для зворотного поширення помилки, тому що ці формули прямо залежать від виду цільової функції.

2. Адаптувати алгоритм бустинга.

Тут розглянемо другий варіант, він менш очевидний. Перший можна залишити в якості вправи.

Змінам піддадуться цільова функція та формування навчальної вибірки для m -го класифікатора.

У якості цільової функції можна використовувати стандартну сумарну квадратичну помилку вихідного сигналу:

$$err = \frac{1}{2} \sum_i (y_i - t_i)^2, \quad (6)$$

у такий спосіб відомі формули для зворотного поширення змінювати немає необхідності.

У формулі (6) не враховуються ваги навчальних прикладів, що не узгоджується з ідеєю бустинга: навчання наступних класифікаторів з упором на приклади, складні для попередніх класифікаторів. Для виправлення цього недоліку можна зробити наступне. Для кожного класифікатора формується своя навчальна вибірка на підставі оновлених ваг прикладів. При цьому, розмір вибірки дорівнює розміру вихідної навчальної вибірки, N , а ймовірність потрапляння i -го прикладу у формовану вибірку рівна:

$$\frac{v_i^{(m+1)}}{\sum_k v_k^{(m+1)}}. \quad (7)$$

Оскільки в початковий момент часу всі ваги прикладів рівні, то для найпершого класифікатора навчальна вибірка буде така ж як і вихідна. Для всіх наступних класифікаторів більш складні для розпізнавання навчальні приклади будуть зустрічатися у вибірці кілька раз, а прості приклади можуть у вибірці бути відсутні. За рахунок цього забезпечується «підвищена увага» до складних навчальних прикладів.

Таким чином в наслідок використання методів ансамбля за рахунок взаємної компенсації недоліків складових методів можна досягти покращену оцінку точності класифікації

Література

1. Гасфилд Д. Строки, деревья и последовательности в алгоритмах. СПб.: Невский диалект, 2003. 654 с.

2. Хайкин С. Нейронные сети: полный курс, 2-е изд., испр. : Пер. с англ. – М. : 000 "И.Д. Вильямс", 2006. – 1104 с.

3. Hengl T. Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning // Nutrient Cycling in Agroecosystems. 2017. № 1 (109). С. 77–102.

Використання методу локалізації при автоматизації діяльності релігійної установи

Ройзнер К.І., Мазурець О.В.

Хмельницький національний університет

Функціональні можливості сучасних інформаційних систем у релігії дозволили сформувати багато мережевих проєктів, в тому числі мережеві богослужіння, сайти сповіді та причастя, віртуальні релігійні бібліотеки тощо. Релігія знаходить своє втілення в інформаційному просторі, віртуальний сегмент все більш помітно впливає на соціально-релігійну дійсність, що сприяє активізації релігійного життя, появі релігійного інформаційного простору, інтернет-церков, поширення віртуального місіонерства, релігійний відеохостінг тощо. З розвитком інформаційних технологій, за допомогою нових способів комунікації, які пропонує Інтернет, розширюються можливості зворотного зв'язку і особистого вибору віруючих, відбувається видозміна релігійних практик і самого інституту релігії, яка стає більш відкритою і більш доступною. Практично всі конфесійні об'єднання організують свої відкриті інформаційні ресурси, де можна не тільки отримати інформацію про організацію, а й взяти участь у якомусь обряді [1].

Релігійні установи використовують інформаційні системи для забезпечення різноманітних комунікаційних та організаційних потреб [2, 3], яке загалом можна розділити на наступні групи:

1. Програмне забезпечення для зовнішнього використання – характеризується наданням широкого відкритого доступу до викладених матеріалів. До цієї категорії належать віртуальні періодичні видання релігійних установ, інформаційні сайти, бібліотеки та сервіси для спілкування.

2. Програмне забезпечення для внутрішнього використання – характеризується накопиченням даних для внутрішнього використання для організаційних, аналітичних цілей та ведення звітності. До цієї категорії належать системи, що автоматизують виконання визначеного обсягу функцій відповідних працівників релігійних установ; зазвичай це прості програмні продукти або офісне програмне забезпечення широкого вжитку.

3. Програмне забезпечення для зовнішньої емуляції внутрішнього використання – характеризується збором даних для внутрішнього використання для організаційних, аналітичних цілей та ведення звітності, яке здійснюється шляхом їх внесення до системи зовнішніми користувачами. До цієї категорії належать системи, що дозволяють прихожанам шляхом використання відкритого зовнішнього доступу вносити свої індивідуальні дані до системи, записуватись на заходи та церемонії, залишати відгуки тощо. Хоча з точки зору даних таке програмне забезпечення можна віднести до категорії програмного забезпечення для внутрішнього використання, з точки зору технологій доступу дане програмне забезпечення можна віднести до категорії програмного забезпечення для зовнішнього використання.

Аналіз інформаційних моделей діяльності релігійних установ різних конфесій та рівня виявив, що значна частина їх елементів має спільні риси, а функціональна модель є подібною. Це визначає можливість уніфікації програмного забезпечення для автоматизації внутрішньої діяльності відповідних релігійних установ.

Метою роботи є використання методу локалізації при автоматизації діяльності релігійної установи в межах відповідної інформаційної технології. Відповідно, особливістю інформаційної технології визначається можливість реалізації програмної системи з можливістю локалізації – налаштування під датові вимоги конкретної релігійної установи незалежно від конфесії та використовуваної мови.

На даному етапі додатки для автоматизації внутрішньої релігійних установ розробляються переважно як цільові програмні продукти, орієнтовані під потреби певної релігійної установи. За такого підходу після етапу розробки (включаючи тестування) слідує етапи налаштування (включаючи впровадження й заповнення БД) та використання (рис. 1).

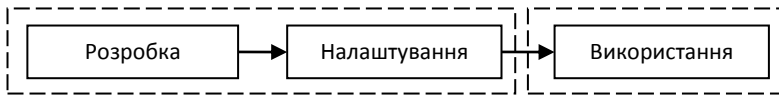


Рисунок 1 – Вихідна послідовність етапів впровадження системи

Характерні риси інформаційних моделей діяльності релігійних установ різних конфесій та рівнів визначають, що значна частина їх елементів має спільні риси, а функціональна модель є подібною. Тому є доцільною уніфікація програмного забезпечення для автоматизації внутрішньої діяльності таких установ. Процес приведення універсальної моделі до цільового вигляду визначається як процес локалізації.

Локалізація програмного забезпечення є його підготовкою до використання в конкретних визначених умовах. Локалізація часто використовується для поширення програмного забезпечення загального призначення в нових географічних чи культурних умовах і включає визначені операції по перевизначенню ряду параметрів інформаційної системи (наприклад, мови, кольорової палітри, ілюстративного матеріалу тощо).

Відповідно до такого підходу, повний цикл переведення системи до експлуатації, характерний для стандартної послідовності, розбивається на ряд етапів, переважна частина яких може бути виконана для цілих категорій релігійних установ. Суттєве зменшення кількості можливих варіантів локалізації на кожному з етапів дозволяє реалізувати їх на рівні, доступному для використання кінцевим користувачем (рис. 2).

Відповідно, кінцевий користувач одержує можливість самостійного впровадження системи для цільового використання.

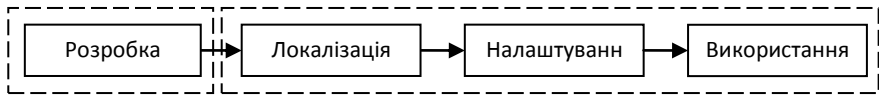


Рисунок 2 – Послідовність етапів впровадження системи з локалізацією

Для визначення даних, що потребують використання у процесі локалізації інформаційної системи, вся множина необхідних модифікацій L_{All} може бути подана у наступному вигляді:

$$(L_{Den} \cup L_{Lang} \cup L_{Org}) \subset L_{All}, \quad (1)$$

де L_{Den} – множина налаштувань по конфесії, L_{Lang} – множина налаштувань по мові, L_{Org} – множина налаштувань по обраній установі.

В (1) до множини налаштувань по конфесії L_{Den} входять налаштування, що визначають спільну для конфесії групу параметрів, які стосуються використання інформаційної системи для визначеної конфесії: множина графічних зображень згідно обраної конфесії, множина кольорів до оформлення інтерфейсу згідно обраної конфесії, множина типів шрифтів інтерфейсу згідно обраної конфесії, множина текстових елементів інтерфейсу згідно обраної конфесії, множина текстових записів у БД згідно обраної конфесії.

До множини налаштувань по мові L_{Lang} входять всі елементи, які визначаються вибором мови інтерфейсу системи: множина мовозалежних графічних зображень згідно обраної конфесії, множина текстових елементів інтерфейсу на обраній мові згідно обраної конфесії, множина текстових записів у БД на обраній мові згідно обраної конфесії.

До множини налаштувань по обраній установі L_{Org} належать безпосередньо дані цільової релігійної установи, що вносяться вручну безпосередньо кінцевими користувачами на перших етапах використання системи, наприклад дані працівників, реквізити установи, плани роботи тощо.

Оскільки метою застосування інформаційної технології є формування елементів інтерфейсу і початкових даних БД, то модифікація системи та її БД визначається як мета застосування підсистеми локалізації (рис. 3). При цьому БД локалізації містить всі елементи всіх розроблених варіантів локалізації. Як результат роботи підсистеми локалізації, проводиться відповідна модифікація елементів підсистеми користувацького інтерфейсу та пере визначення визначених записів в користувацькій базі даних.

Відповідно, інформаційна технологія автоматизації діяльності релігійної установи передбачає проведення локалізації як результат ряду послідовних етапів (рис. 4).

Найбільш загальним є вибір актуальної конфесії (Блок 1), що визначає область подальших дій. У залежності від вибору актуальної конфесії, проводиться обробка елементів множини налаштувань по конфесії (Блок 2),

внаслідок чого визначаються елементи, які будуть використані для модифікації. Вибір мови інтерфейсу системи (Блок 3) остаточно формує вибірку елементів, що будуть використані при локалізації за критерієм вибору мови (Блок 4). Внесення елементів множини налаштувань по обраній установі (Блок 5) полягає у занесенні вручну до бази даних початкових даних цільової релігійної установи кінцевими користувачами.

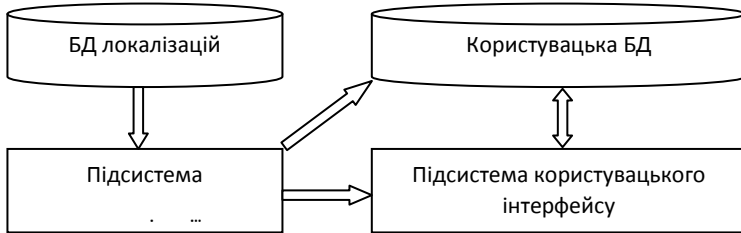


Рисунок 3 – Потіки даних при локалізації системи

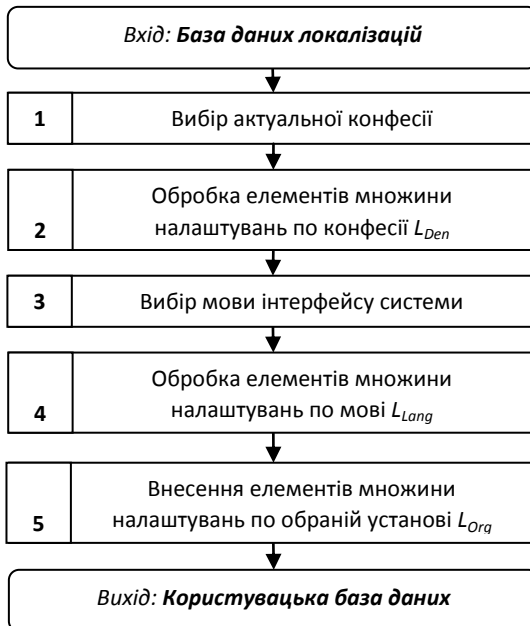


Рисунок 4 – Загальна схема інформаційної технології автоматизації діяльності релігійної установи з локалізацією

В результаті формується користувацька база даних, яка й використовується при роботі кінцевих користувачів з інформаційною системою автоматизації діяльності релігійної установи.

Таким чином, в рамках роботи було викладено узагальнену математичну модель локалізації системи автоматизації діяльності релігійної установи. Розглянуто інформаційну технологію автоматизації діяльності релігійної установи.

Особливостями розглянутої інформаційної технології автоматизації діяльності релігійної установи є:

- можливість локалізації – модифікації програмної системи з можливістю налаштування під датові вимоги конкретної релігійної установи незалежно від конфесії та використовуваної мови;
- простота локалізації – можливість проведення всіх локалізацій кінцевим користувачем перед початком роботи з системою;
- остаточність локалізації – виконані в процесі локалізації модифікації використовуються в подальшому для виконання функцій системи й не можуть бути перевизначені.

Отже, запропонована інформаційна технологія автоматизації діяльності релігійної установи дозволяє розробляти кросконфесійні інформаційні системи автоматизації діяльності релігійної установи, які можуть бути легко модифіковані для автоматизації внутрішньої діяльності конкретних релігійних установ різних конфесій та на різних мовах інтерфейсу. До прикладних особливостей використання розглянутого методу локалізації при автоматизації діяльності релігійної установи належать можливість, простота та остаточність локалізації.

Література

1. Пронько А. М. Релігія в епоху інформаційних технологій / А. М. Пронько // Сучасні тенденції розвитку науки. – Харків, 2017. – С. 161–163.
2. Karaflogka A. Religious Discourse and Cyberspace / A. Karaflogka // Religion. – 2002. – № 32. – P. 279-291.
3. Зеленська Л. І. Картографічне забезпечення релігійного туризму (регіональний компонент) / Л. І. Зеленська, Д. Є. Котлярчук // Національне картографування: стан, проблеми та перспективи розвитку: Збірник наук. праць. – К., 2005. – Вип. 2. – С. 158–166.

Інформаційна технологія рекурсивного семантичного аналізу текстів шляхом дисперсійного оцінювання слів

Сергієва О.О., Мазурець О.В.

Хмельницький національний університет

Семантичний аналіз тексту є етапом у послідовності дій алгоритмів автоматичного розуміння текстів, що полягають у виділенні семантичних відношень, формуванні семантичного подання текстів. Один з можливих варіантів відображення семантичного подання – структура, що складається з

текстових елементів. Глибина семантичного аналізу може бути різною, а в існуючих системах найчастіше будується тільки синтаксико-семантичне подання тексту або окремих фрагментів, до яких відносять анотації, реферати та переліки ключових слів [1].

Перелік ключових слів тексту є найбільш семантично стиснутим результатом семантичного аналізу тексту, й пошук ефективних методів автоматизації його формування відкриє надає можливість вирішення багатьох похідних задач.

Метою роботи є розробка інформаційної технології семантичного аналізу текстів шляхом дисперсійного оцінювання слів.

Схему відповідної інформаційної технології подано на рисунку 1. В якості метода семантичного аналізу текстів використовується метод рекурсивного дисперсійного оцінювання [2] для пошуку ключових термінів у текстовому контенті цифрового файлу .docx.

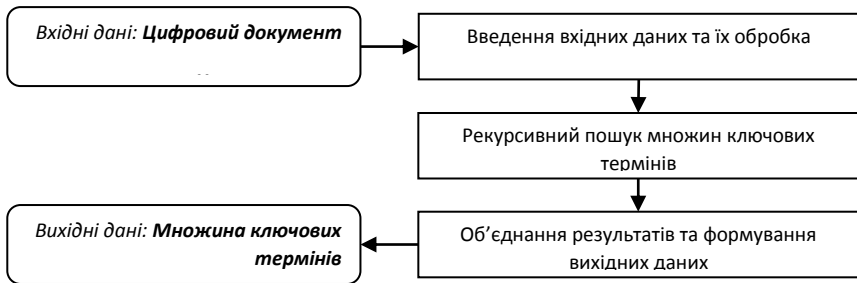


Рисунок 1 – Схема інформаційної технології рекурсивного семантичного аналізу текстів шляхом дисперсійного оцінювання слів

Наведені на рисунку 1 основні етапи автоматизованого аналізу текстового контенту цифрових документів передбачають на першому етапі одержання системою вхідних даних, до яких належать власне цифровий документ, що аналізується, та необов'язкові параметри роботи, якщо існує необхідність їх перевизначення. Після чого проводиться рекурсивний пошук множин ключових термінів (рис. 2), по одній ітерації для кожного варіанту розмірності термінів (1- n слів у терміні). На завершальному етапі роботи системи виконується об'єднання результатів окремих ітерацій пошуку множин ключових термінів, за результатами чого впорядкована й обмежена результуюча множина ключових термінів виводиться в якості вихідних даних.

Вхідними даними інформаційної технології рекурсивного аналізу текстів, є цифровий документ (файл з розширенням .docx), контент якого містить текст для аналізу; та набір необов'язкових параметрів роботи системи: максимальна кількість слів у терміні n та параметр щільності

ключових термінів у тексті P . По замовчуванню використовуються параметри $n=5$ та $P=15$. Вихідними даними інформаційної технології є множина ключових термінів тексту.

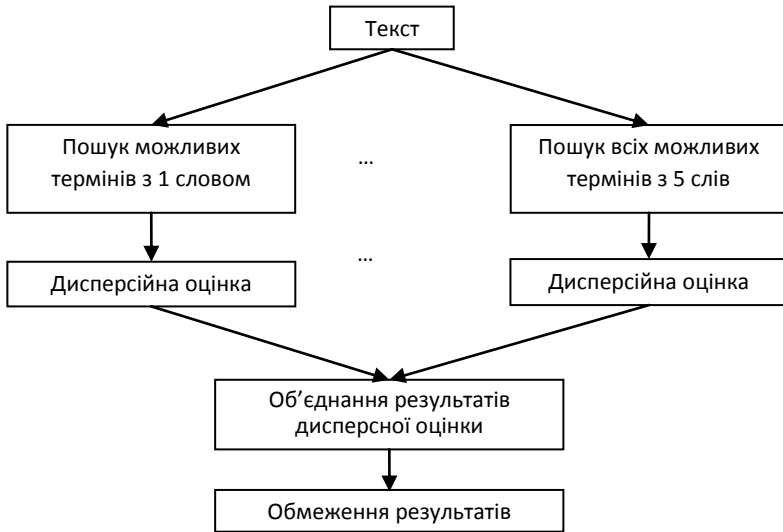


Рисунок 2 – Схема рекурсивної функції методу дисперсійного оцінювання слів в інформаційній технології

Дисперсійний аналіз, що використовується для пошуку ключових термінів, є статистичним методом оцінки зв'язку між факторними й результативними ознаками в різних групах, відібраний випадковим чином, заснований на визначенні розходжень значень ознак. В основі дисперсійного аналізу лежить аналіз відхилень всіх одиниць досліджуваної сукупності від середнього арифметичного. Як міра відхилень береться дисперсія – середній квадрат відхилень. Відхилення, викликані впливом факторної ознаки порівнюються з величиною відхилень, викликаних випадковими обставинами. Якщо відхилення, викликані факторною ознакою, більш істотні, ніж випадкові відхилення, то вважається, що фактор впливає на результуючу ознаку. Таким чином, дисперсійна оцінка є оцінкою дискримінантної сили слів й дозволяє відділити із загальної множини швидкоживаних у тексті слів слова, що розташовані рівномірно.

На основі запропонованого підходу було створено тестову автоматизовану систему рекурсивного семантичного аналізу текстів шляхом дисперсійного оцінювання слів, яка продемонструвала достатньо високу ефективність, особливо для великих текстів. Розроблена система дозволяє за вхідними даними у вигляді завантаженого файлу цифрового документу з

розширенням .docx, контент якого містить текст для аналізу (рис. 3) та наборок параметрів роботи системи одержати множини ключових термінів тексту (рис. 4).

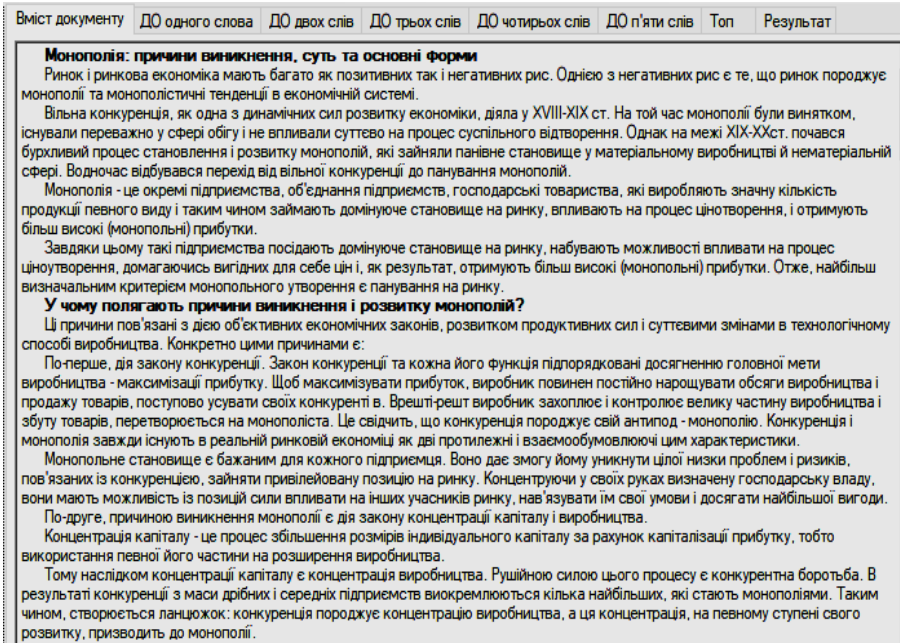


Рисунок 3 – Завантажений у систему файл цифрового документу для аналізу

Розроблена система рекурсивного семантичного аналізу текстів виконує такі функції:

- зчитування .docx файлу та витяг його текстового контенту;
- побудова дерева структури документу;
- вибір елемента дерева структури документу для аналізу;
- формування множини термінів тексту;
- оцінка важливості термінів тексту методом дисперсійного оцінювання;
- сортування й обмеження результуючої множини термінів;
- збереження результатів пошуку в базу даних та перегляд збережених результатів.

Отже, було запропоновано інформаційну технологію семантичного аналізу текстів шляхом дисперсійного оцінювання слів. На основі запропонованої інформаційної технології було створено тестову

автоматизовану систему рекурсивного семантичного аналізу текстів шляхом дисперсійного оцінювання слів, яка продемонструвала достатньо високу ефективність, особливо для великих текстів. Подальші дослідження спрямовані на поширення можливостей технології на ефективну роботу із складними реченнями та семантичними конструкціями.

Вміст документу	ДО одного слова	ДО двох слів	ДО трьох слів	ДО чотирьох слів	ДО п'яти слів	Топ	Результат
	№	Терміни	DE	Кількість			
▶	0	капіталу	1,70698654098086	14			
	1	високі	1,55724218127515	5			
	2	товарів	1,54268695191629	6			
	3	монополій	1,45758389717667	15			
	4	процес	1,41580159202711	8			
	5	конкуренція	1,39343668715595	24			
	6	ціни	1,39114967288911	8			
	7	є	1,32625783518659	20			
	8	економіці	1,31735581056624	4			
	9	панування	1,30442804610273	4			
	10	підприємств	1,30074249252838	10			
	11	між	1,29162016769849	10			
	12	яка	1,25010128495193	5			
	13	виробництва	1,23439179577515	23			
	14	цін	1,23258084468753	9			
	15	цьому	1,2262721430292	4			
	16	в	1,17654363812124	32			

Рисунок 4 – Результат визначення множини ключових термінів тексту

Перевагою даної інформаційної технології рекурсивного семантичного аналізу текстів шляхом дисперсійного оцінювання слів є те, що вона дозволяє проводити пошук ключових термінів без використання спеціальних цифрових словників чи баз даних корпусу слів відповідної мови.

Література

1. Сергієва О. О., Мазурець О. В. Інтелектуальна система автоматизованого стиснення текстів / О. О. Сергієва, О. В. Мазурець // Матеріали VI Міжнародної науково-практичної конференції «Інформаційні управл. системи та технології ICST-ODESSA-2017». Одеса – 2017. – С.283-285.
2. Бармак О. В., Мазурець О. В. Методи автоматизації визначення семантичних термінів у навчальних матеріалах / О. В. Бармак, О. В. Мазурець // Вісник Хмельницького національного університету. Сер.: Технічні науки. Хмельницький. – 2015, №2(223). – С.209-213.

Інформаційна модель автоматизації супроводу навчального процесу

Ставнійчук М.В., Мазурець О.В.

Хмельницький національний університет

Навчальний процес у вищих навчальних закладах здійснюється у різноманітних формах, таких як навчальні заняття [1], виконання індивідуальних завдань, самостійна робота студентів, практична підготовка, контрольні заходи тощо. Крім цього, з окремими студентами проводиться індивідуальні навчальні заняття з метою підвищення рівня їх підготовки та розкриття індивідуальних творчих здібностей. Індивідуальні навчальні заняття організуються за окремим графіком з урахуванням індивідуального навчального плану студента і можуть охоплювати частину або повний обсяг занять з однієї або декількох навчальних дисциплін, а в окремих випадках - повний обсяг навчальних занять для конкретного освітнього або кваліфікаційного рівня [2].

Щоб полегшити комунікацію між викладачами та студентами в рамках навчального процесу та забезпечити більш ефективну координацію й планування елементів навчальних заходів, є доцільним використання спеціалізованої автоматизованої системи супроводу навчального процесу.

Метою роботи є розробка інформаційної моделі автоматизації супроводу навчального процесу, яка дозволяє створення програмної системи для координації та планування елементів навчальних заходів. Задачами, які може дозволити вирішити така система, є інтерактивне інформування викладачами про їх доступність на робочих місцях, перегляд оголошень викладачів з планування їх навчальних заходів, онлайн-запис на консультації та інші навчальні заходи тощо.

За результатами проведеного аналізу інформаційного забезпечення предметної області, визначено такі групи функцій автоматизованої системи супроводу навчального процесу (рисунок 1): робота з відображенням присутності викладачів, розміщення і перегляд оголошень викладачів, розміщення коментарів студентів до оголошень викладачів, користування чатами та адміністрування системи.

Група функцій «Робота з відображенням присутності викладачів» забезпечує відображення присутності викладачем в університеті, зміни статусу присутності в режимі онлайн, а також дає можливість студентам в режимі онлайн переглядати дані про присутність викладачів в університеті. Група функцій «Розміщення та перегляд оголошень викладачів» містить в собі функціональні можливості по створенню оголошень викладачами для студентів. Група функцій «Робота з коментарями до оголошень викладача» поєднує в собі можливості по створенню коментаря до обраного повідомлення й перегляд всіх коментарів до конкретного повідомлення. Група функцій «Користування чатами» включає в себе можливості по

обміну повідомленнями всередині груп користувачів – викладачів та студентів. Група функцій «Адміністрування системи» включає в себе можливості по створенню нових користувачів, додаванню нових таблиць, видаленню непотрібної інформації, додаванню нових адміністраторів системи, перегляду статистичних даних про роботу системи, а також повне редагування всіх доступних таблиць БД.

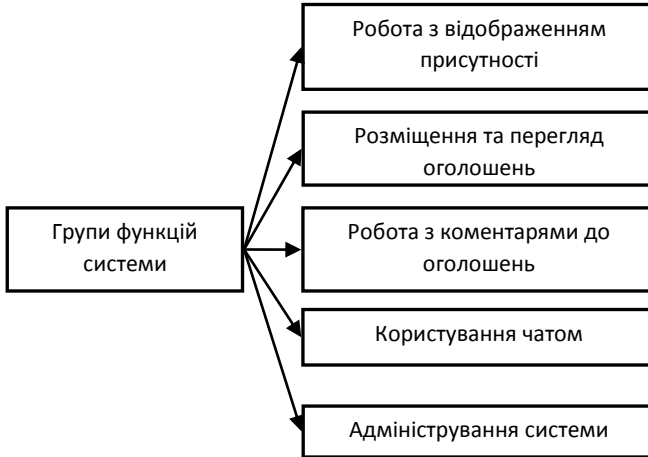


Рисунок 1 – Схема груп функцій системи

Інформаційна модель автоматизації супроводу навчального процесу передбачає автоматизацію виконання ряду функцій. Відповідно, до загальної сукупності функцій F відносяться такі групи функцій:

$$F = W_1 + W_2 + W_3 + W_4 + W_5, \quad (1)$$

де F – автоматизована система супроводу навчального процесу; W_1 – функції роботи з відображенням присутності; W_2 – функції розміщення та перегляд оголошень; W_3 – функції роботи з коментарями до оголошень; W_4 – функції користування чатом; W_5 – функції адміністрування системи.

До функцій роботи з відображенням присутності W_1 відносяться відповідні функції:

$$W_1 = \langle A, B, C \rangle \quad (2)$$

де A – функція відмітки про власну присутність в університеті, є обов’язковим атрибутом; B – функція відмітки про власну відсутність в університеті, є обов’язковим атрибутом; C – функція перегляду поточних присутностей викладачів в університеті, є необов’язковим атрибутом.

До функцій розміщення та перегляд оголошень W_2 відносяться відповідні функції:

$$W_2 = \langle D, E, F, G, H, I \rangle \quad (3)$$

де D – функція створення викладачем нового оголошення, є необов’язковим

атрибутом; E – функція перегляду переліку власних оголошень автором, є необов’язковим атрибутом; F – функція редагування власних оголошень автором, є необов’язковим атрибутом; G – функція видалення власних оголошень автором, є необов’язковим атрибутом; H – функція перегляду користувачем актуальних оголошень обраного викладача, є необов’язковим атрибутом; I – функція автоматичного переведення оголошень до не актуального стану при виході терміну актуальності, є обов’язковим атрибутом.

До функцій роботи з коментарями до оголошень W_3 відносяться відповідні функції:

$$W_3 = \langle J, K, L \rangle \quad (4)$$

де J – функція додавання нового коментаря до обраного оголошення, є необов’язковим атрибутом; K – функція перегляду переліку коментарів до обраного оголошення, є необов’язковим атрибутом; L – функція видалення обраного власного коментаря до оголошення, є необов’язковим атрибутом.

До функцій користування чатом W_4 відносяться відповідні функції:

$$W_4 = \langle M, N, O \rangle \quad (5)$$

де M – функція перегляду переліку доступних чатів, є обов’язковим атрибутом; N – функція створення нових повідомлень у вибраному доступному чаті, є необов’язковим атрибутом; O – функція перегляду треду (переліку повідомлень) обраного доступного чату, є обов’язковим атрибутом.

До функцій адміністрування системи W_5 відносяться відповідні функції:

$$W_5 = \langle P, Q, R, S, T, U, V, X, Y \rangle \quad (6)$$

де P – функція додавання адміністраторів, є необов’язковим атрибутом; Q – функція перегляду статистики сайту, є необов’язковим атрибутом; R – функція видалення адміністраторів, є необов’язковим атрибутом; S – функція додавання нових користувачів, є обов’язковим атрибутом; T – функція видалення користувачів, є обов’язковим атрибутом; U – функція додавання груп, є обов’язковим атрибутом; V – функція додавання наукових ступенів, є обов’язковим атрибутом; X – функція редагування груп, є обов’язковим атрибутом; Y – функція редагування наукових ступенів, є обов’язковим атрибутом.

За наведеною інформаційною моделлю було розроблено тестову систему автоматизації супроводу навчального процесу на платформі PHP [3]. Для формування архітектури системи було створено відповідну структуру контролерів (Рисунок 2). Клас `PlanController` містить методи `add()`, `history()`, `save()`, `edit()`, `update()`, `remove()`, `plansByTeacher()`. Дані методи призначені для перегляду всіх оголошень студентом, формування переліку оголошень по вибраному викладачеві; створення, редагування, видалення оголошень тощо. Клас `TeacherController` містить методи `edit()`, `update()`, `changeStatus()`, `getTeacher()`. Дані методи призначені для редагування особистих даних викладача, зміни статусу присутності, а також формування списку викладачів по вибраній кафедрі. Клас `StudentController` містить методи `edit()`, `update()`,

getStudent(). Дані методи призначені для редагування особистих даних студента, а також для формування списку студентів по обраній групі. Решта контролерів мають аналогічну структуру. Запропонована архітектура є достатньою для реалізації всіх функцій автоматизованої системи супроводу навчального процесу.

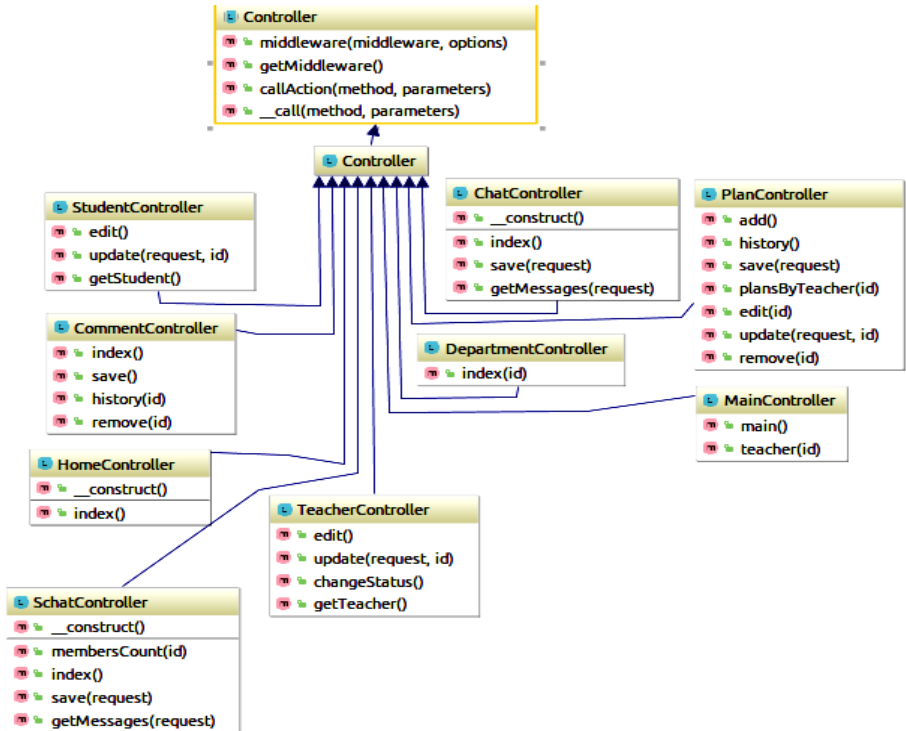


Рисунок 2 – Структура контролерів та методів в контролерах системи автоматизації супроводу навчального процесу

Для забезпечення роботи всіх поставлених функцій і завдань, а також для забезпечення функціонування системи було розроблені програмні модулі відповідні до перерахованих вище функцій. На рисунку 3 зображено головну сторінку системи, яка має стислий опис та початкову інформацію про систему. З неї відкривається доступ до основних функцій системи.

За результатами тестування всіх компонентів програми було зроблено висновок, що розроблена система працездатна та працює без збоїв. Таким чином, запропонована інформаційна модель автоматизації супроводу навчального процесу є коректною та відповідною поставленій меті.

Отже, в результаті виконаної роботи було розроблено інформаційну

модель автоматизації супроводу навчального процесу, яка дозволяє створення програмної системи для координації та планування елементів навчальних заходів. Задачами, які може дозволити вирішити така система, є інтерактивне інформування викладачами про їх доступність на робочих місцях, перегляд оголошень викладачів з планування їх навчальних заходів, онлайн-запис на консультації та інші навчальні заходи тощо.

Автоматизація виконання наведених функцій за допомогою відповідного програмного забезпечення дозволить полегшити комунікацію між викладачами та студентами в рамках навчального процесу та забезпечити більш ефективну координацію й планування елементів навчальних заходів

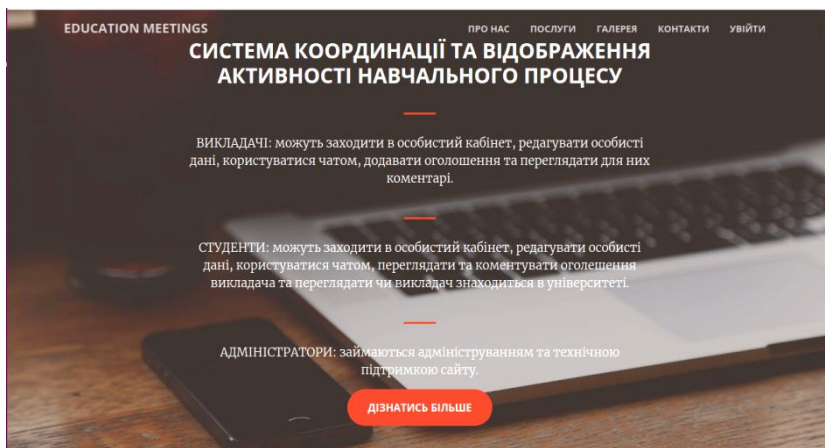


Рисунок 3 – Стартова сторінка системи автоматизації супроводу навчального процесу

Література

1. Демчук М. В., Мазурець О. В. Автоматизована система ведення розкладу занять у вузі / М. В. Демчук, О. В. Мазурець // Збірник наукових праць за матеріалами десятої міжнародної науково-технічної конференції «Актуальні проблеми комп'ютерних технологій 2016». Хмельницький – 2016. – С.65-71.
2. Організація навчального процесу у ВНЗ [Електронний ресурс]. – Режим доступу: <http://education.nuos.edu.ua>
3. Ставнійчук М. В., Мазурець О. В. Автоматизована інтерактивна система супроводу навчального процесу / М. В. Ставнійчук, О. В. Мазурець // Матеріали VII Міжнародної науково-практичної конференції «Інформаційні управляючі системи та технології ICST-ODESSA-2018». Одеса – 2018. – С.80-83.

Інформаційна технологія класифікації текстової інформації

Терлецький Ю.В, Свірневський М.С., Скрипник Т.К.

Хмельницький національний університет

В наш час завдяки високому розвитку технологій більша частина сучасних великих компаній уже здійснила перехід на ведення безпаперового документообігу. Широке впровадження одержали системи, що використовують концепцію та технології ІІМ (системи, що забезпечують зберігання даних і оптимальний час доступу до них) [1], а також PDM-системи [2]. Система PDM управляє обміном даними про виріб, забезпечує взаємодію з будь-якими корпоративними додатками та відслідковує внесення змін у виріб і документацію про нього. Крім того системи, що існують, бізнесаналітики здійснюють інтелектуальну обробку збережених даних. Проте, не створене ефективних систем, що обробляють текстову складову інформації про виріб.

Однак очевидно, що внаслідок постійного збільшення обсягу збережених електронних даних, найважливішим завданням є організація ефективного пошуку усередині документації компанії. Так, наприклад, починаючи розробку нового продукту, розроблювачі повинні провести аналіз існуючих розв'язків. На великому підприємстві з великою історією виробництва в першу чергу вивчаються власні архіви.

Пошукові машини, що здійснюють пошук по наявній документарній базі, як правило, надають можливість формулювання запиту за допомогою ключових слів. Це ставить перед користувачем ряд проблем.

Найбільш уразливі в цьому випадку однослівні терміни. На відміну від них, багатослівні конструкції мають більшу стійкість. У пропонованому підході виходили з міркування, що, незважаючи на певний ступінь зміни термінологічної лексики та системи понять у більш пізній документації, для всіх релевантних документів у цілому зберігається велика множина загальних слів.

Відповідно до цього, пропонується застосувати наступний підхід: проводити пошук не по ключових словах, а по характерній лексиці. Для цього вибирається т.зв. документ-зразок, з якого за допомогою викладених нижче принципів виділяються неоднослівні комбінації, які надалі використовуються для прецедентного пошуку усередині документації. Оскільки маємо справу з технічною документацією, то в якості документа-зразка використовується технічне завдання (ТЗ). Технічне завдання містить у собі короткий, але повний опис постановки завдання, а також методів її розв'язку. Таким чином, у ньому обов'язково є присутнім уся основна термінологія.

Ціль - це пошук документів, які могли б бути корисні в поточному проекті, з використанням ТЗ як запиту. Для її розв'язку потрібно розробити метод, який дозволить би обчислювати подібності документів. На основі

використання цього підходу можна проводити ранжування документів і подавати на вихід найбільш релевантні. Даний підхід повинен враховувати і багатослівні конструкції, що зустрічаються в тексті документації.

Друге завдання - це перевірка повноти підсумкової документації по поставленому ТЗ. Для цього необхідно автоматично виділяти з тексту ТЗ вихідні дані та необхідні результати для розроблювального продукту. На основі цієї інформації проводиться автоматична перевірка документації. Наскільки повні вимоги, зазначені в ТЗ, фігурують у підсумковому звіті та у яких місцях документації наведений їхній опис. Подібним чином можемо оцінити ступінь повноти документації у відсотках від поставлених завдань. При зниженні отриманого значення нижче граничного, документація вважається неповною, і маємо можливість вказати, які саме питання із ТЗ відсутні в підсумковій документації. В останньому випадку виконавець повинен буде переробити документацію та, можливо, переглянути свою роботу.

З погляду споживача, система повинна проводити попереднє дослідження підсумкової документації на предмет явної відсутності опису деяких питань, описаних у ТЗ. Список вимог може коректуватися експертом вручну. Система не передбачає явної протидії їй, коли розроблювачі включають у підсумкову документацію «шум», що імітує фрагмент звіту. Таким чином, негативна відповідь системи повина розглядатися як привід для направлення документації на узгодження або доробку. Позитивна відповідь системи служить сигналом для її перевірки стороною, що ухвалює, тому що гарантує лише наявність інформації у звіті, але не її глибину.

У наш час розроблено чимала кількість різних методів, використовуваних для розв'язку подібних завдань. Розглянемо деякі з них, розподілені по трьом групам.

Пошук по ключових словах і словосполучення може проводитися різними методами. Традиційним є використання статистичного заходу $tf*idf$, що показує «інформативність терміна» [3]. Tf (term frequency) являє собою відношення числа входжень деякого терміна до безлічі всіх слів у документі. Idf (inverse document frequency) - інверсія частоти, з якої те ж слово зустрічається у всіх документах розглянутої колекції. У цілому цей підхід враховує ступінь важливості терміна усередині одного документа, знижуючи вагу для загальноживаної лексики (що має високу частотність у всіх документах розглянутої колекції). Термін з високою зустрічальністю вважається малоінформативним, тому що, швидше за все, є елементом стилістики. У той же час, що рідко зустрічається термін у рамках даного підходу вважається елементом шуму та одержує низьке значення вагової характеристики.

Для словосполучень поряд із підходом $tf*idf$ використовується метод виділення колокацій (невипадкових комбінацій двох і більш лексичні

одиниць), наприклад, підходу MI, t-score або log-score. З їхньою допомогою може бути отриманий набір неоднослівних комбінацій, характерних для розглянутої предметної області. Для цього, як показали дослідження, описані в [4], краще підходить підхід MI. Підхід t-score дозволяє скоріше виявити стилістичні особливості тексту у рамках нашого завдання може служити скоріше для виявлення термінів, що підлягають відсіванню.

Методи, засновані на моделі векторного простору, використовують формований заздалегідь вектор ознак документа, що включає в себе список найбільш значимих термінів. Подальше порівняння документів проводиться шляхом оцінки подібності між отриманими векторами з використанням різних заходів подібності. У даній роботі розглянемо часто застосовувані коефіцієнт Дайса та косинусну захід [1, 2].

Для формування вектора ознак документа зазвичай використовується деяка підмножина слів. Крім уже згаданого підходу $tf*idf$ можуть використовуватися, наприклад «опорні слова». З множини слів усіх документів колекції вибирається N «опорних» і для кожного документа формується N -мірний вектор ознак. У вектор заноситься 1, якщо частота даного слова в документі перевищує деяке граничне значення. Подібними вважаються документи в яких вектори ознак на даний момент подібні, та використовують невеликі фрагменти документів для їхнього порівняння або їх оцінки. Так, наприклад, група алгоритмів, що базуються на алгоритмі випадкових поліномів Коропа-Рабина, опирається на виявлення подібних груп, послідовних слів довжини k (шинглів) або дактилограмм - підстрок документа фіксованої довжини. Для документа обчислюється фіксовану кількість шинглів або дактилограмм, після чого проводиться порівняння подібних послідовностей. Для більш швидких методів застосовується обчислення хеш-функції для отриманих послідовностей. Документи вважаються подібними, якщо отримане значення хеш-функції збігається. Більш повільні, але більш точні методи порівнюють отримані послідовності та на підставі оцінки їх збігу робиться висновок про ступінь і ймовірності подібності документів.

Підходи використовуються для визначення ступеня подібності документів між собою. У якості запиту подається документ, для якого здійснюємо пошук по подібності. Обчислені значення дозволяють ранжувати документи зі сховища по релевантності. Очевидно, що підходи будуть давати різні результати.

На основі емпіричного аналізу текстів технічної документації, що містить ТЗ або постановку завдання, було помічене, що досить розповсюдженими в тексті є дієслівні конструкції певного виду. Передбачається, що усередині пропозицій з такими конструкціями утримується інформація, що стосується вимог, зазначених у ТЗ, і/або опис їх реалізації.

Відповідно до даного припущення порівняння документації

проводиться в такий спосіб. Заздалегідь визначається список дієслівних конструкцій, характерних для технічних текстів і текстів обраної предметної області, що та описують вимоги до продукту. З наявних конструкцій формується список шаблонів. Далі здійснюється пошук відібраних конструкцій у тексті документа-зразка (у нашому випадку це ТЗ). Оскільки цікавить лексика, що характеризує даний документ, то для подальшого аналізу з тексту вибирається відрізок, більший, ніж пропозиція, що безпосередньо містить знайдену конструкцію. Визначення розміру «вікна пошуку» являє собою окреме завдання. У першому наближенні було вирішено використовувати для подальшої обробки інтервал, що включає, крім знайденого пропозиції, ще два сусідні з кожної сторони. У середині цього інтервалу з найбільшою ймовірністю буде зустрічатися тематично значима лексика, оскільки важко припустити, що опис завдання / вимог буде виражено в одній пропозиції.

Виходячи з перерахованих вище міркувань, була реалізована програма, що дозволяє говорити про наступні проміжні результати. На підставі лексики, яка береться з тексту технічного завдання з використанням методики шаблонів, по всіх текстах колекції підсумкової документації шукаються відповідності фрагментам із ТЗ. Знайдені відповідності рівняються з вихідними за допомогою заходів подібності. Числові характеристики, які одержали по цих заходах, дозволяють зробити однозначний висновок про тематичну зв'язаність вихідного ТЗ і відповідної йому колекції документації.

Крім того, було проведено ранжирування документів по кожному з використовуваних підходів (NSL, SSL, а також модифікованої косинусної міри та коефіцієнту Дайса). Це робилося для того, щоб переконатися в тому, що розглянутий метод не дає високої оцінки нерелевантним документам. Системі було запропоновано зрівняти вихідний документ (ТЗ) з колекцією інших. У колекції, крім документів, що відповідають даному ТЗ, були представлені свідомо нерелевантні документи (як із близьких, так і з незв'язаних предметних областей). Експерименти показали, що в кожному випадку найвищий ступінь подібності одержували документи, що відповідають ТЗ, причому числові характеристики, одержувані по підходах, відрізнялися, у випадку релевантних документів, на порядок.

Ці результати підтверджують доцільність подальшого розвитку обраного підходу. Як наслідок, потрібне вирішення цілого ряду проблем, пов'язаних із забезпеченням якості роботи методу.

Література

1. Гасфилд Д. Строки, деревья и последовательности в алгоритмах. СПб.: Невский диалект, 2003. 654 с.
2. Елкин С.В., Клышинский Э.С., Стеклянный С. Е., Проблемы создания универсального морфосемантического словаря // Сб. трудов

Международных конференций IEEE AIS'03 и CAD-2003, том 1, Дивноморское. 2003. стр. 159163.

3. Захаров В.П., Хохлова М.В. Анализ эффективности статистических методов выявления коллокаций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии: по материалам ежегодной Международной конференции «Диалог» (2010). Вып. 9 (16), сс. 137-143

4. Пивоварова Л., Ягунова Е. Извлечение и классификация терминологических коллокаций на материале лингвистических научных текстов // Терминология и знание: Материалы II Международного Симпозиума (Москва, 21-22 мая 2010г.) - М., 2010.

Інформаційна технологія автоматизованого випадково генерування корпусів відповідно до гаусівського розподілу

Тимчак А.В., Лищук О.А., Скрипник Т.К.

Хмельницький національний університет

Основні методи досліджень, використовувані в даній роботі, це методи теорії ймовірностей і математичної статистики, методи комп'ютерного моделювання та методи оптимізації.

Актуальність даної теми полягає в тому, що оцінки параметрів моделі, отримані за допомогою рангового методу, у деяких випадках є більш точними, чим МНК-оцінка. Так, наприклад, викиди в даних меншою мірою впливають на оцінки параметрів моделі, побудованої за допомогою рангового методу, чим на МНК-оцінка. Причиною цього є те, що функція втрат для МНК-оцінки містить у собі квадрати відхилень спостережуваних значень залежної величини від її оцінок у рамках моделі, у той час як у функцію втрат рангової моделі ці відхилення входять лінійно. Також варто відзначити випадки, коли шуми в моделях мають «двогорбий розподіл» – розподіл з функцією щільності, що має дві вершини максимуму. Тоді ефективність МНК-оцінки параметрів регресії буде нижче в порівнянні з оцінками отриманими ранговим методом.

У рамках цієї роботи планується виконати наступні завдання:

- розробити та чисельно реалізувати алгоритм побудови рангової оцінки невідомих параметрів регресії;
- змоделювати регресійні залежності з погрішностями, що мають розподіл з «важкими» хвостами;
- провести чисельний порівняльний аналіз рангових оцінок із МНК-оцінками;
- обчислити аналітично асимптотичну відносну ефективність рангового методу стосовно МНК при різних розподілах помилок;
- провести експеримент у порівнянні стійкості рангової МНК-оцінки параметрів моделі до викидів у реальних даних.

Мета досліджень: зробити висновки про застосовність рангового методу в завданні оцінювання параметрів і сформулювати конкретні рекомендації із застосування одного з розглянутих методів при різних розподілах шумів на розподілах Гауса.

Отже, у рамках даної роботи розглядаються такі розподіли випадкових величин, як розподіли Гаусса та Лапласа, трикутний розподіл (розподіл Сімпсона) і «двогорбі» розподіли, які модулюються на основі гаусівських і трикутних розподілів. Крім цього розглядаються розподіл Коші, розподіл Стьюдента з невеликим числом ступенів свободи, розподіл Тьюкі та логістичний розподіл. Останні розподіли відносяться до розподілів з «важкими хвостами».

Розподіл Гаусса з дисперсією $\sigma^2 > 0$ і математичним очікуванням m має функцію щільності $f = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$. У проведених експериментах

гаусівська випадкова величина з математичним очікуванням $m=0$ і дисперсією $\sigma^2=1$ моделюється за допомогою стандартних процедур. На рисунку. 1 зображена функція щільності цієї величини.

Розподіл Лапласа із асиметрією β і коефіцієнтом масштабу $\alpha > 0$ має щільність $f = \frac{\alpha}{2} e^{-\alpha|x-\beta|}$. В експериментах розглядається величина з $\beta=0$ і $\alpha=1$.

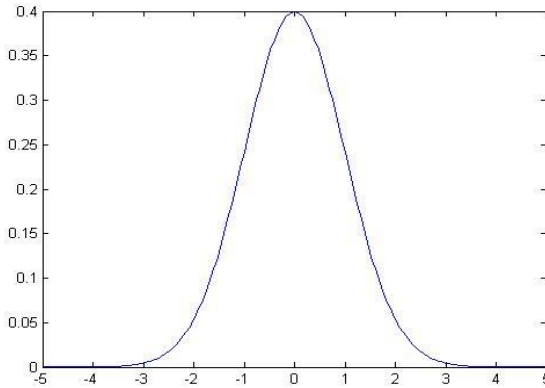


Рисунок 1 – Графік щільності розподілу Гаусса

Величина з таким розподілом моделюється як різниця двох величин з експонентним розподілом. Кожна із цих величин, у свою чергу, обчислюється як логарифм рівномірно розподіленої на відріжку від 0 до 1 випадкової величини, поділеної на $-\alpha$: $z = \frac{\ln(u)}{-\alpha}$.

Для проведення порівняльного аналізу потрібно побудувати регресійні залежності. В експериментах розглядаються моделі з $n = 50$ спостереженнями та $(m+1) = 3$ параметрами, включаючи вільний член. Дані генеруються в такий спосіб:

- спочатку випадковим чином генерується матриця X з даними, рівномірно розподіленими на деякому відрізку. Матриця має n рядків і m стовпців.

- до матриці X приписується стовпець із одиниць для того, щоб будувати моделі з вільним членом.

- згідно з розглянутим розподілом генерується n -мірний вектор-стовпець помилок ε .

- задається m -мірний вектор-стовпець θ з реальними значеннями параметрів лінійної регресійної моделі.

- здується вектор спостережень $Y = X\theta + \varepsilon$

Для кожної побудованої регресійної залежності в експериментах обчислюються рангова МНК-оцінка вектора її параметрів.

Згідно Дж. Себеру МНК-оцінка вектора параметрів лінійної регресійної моделі має вигляд

$$\hat{\theta}_{\text{МНК}} = (X^T X)^{-1} X^T \theta \quad (1)$$

Для побудови наближеної МНК-оцінки використовується метод симплексного пошуку з метою мінімізації функції втрат

$$S = \sum_{k=1}^n |y_k - \theta_0 - x_{k1}\theta_1 - \dots - x_{km}\theta_m| \quad (2)$$

Для побудови наближеної рангової оцінки тим же методом проводиться мінімізація функції втрат D .

Оскільки дані генеруються випадково, то слід уникати впливу якогось конкретного розсіювання даних на загальну картину. Тому для того самого значення заданих параметрів дані генеруються 100 раз у проведених експериментах, щораз на основі цих даних будуються оцінки параметрів, обчислюються значення критерію якості оцінок, і потім визначається вибіркова середня якості оцінок параметрів регресійних моделей для кожного методу. Критерієм якості оцінки в цьому випадку буде виступати сума квадратів різниць дійсного значення параметра і його оцінки

$$d^2(\hat{\theta}, \theta) = \sum_{i=0}^m (\hat{\theta}_i - \theta_i)^2,$$

де $\theta = \begin{pmatrix} \theta_0 \\ \dots \\ \theta_m \end{pmatrix}$ – вектор з дійсними значеннями параметрів, $\hat{\theta} = \begin{pmatrix} \hat{\theta}_0 \\ \dots \\ \hat{\theta}_m \end{pmatrix}$ – вектор з оцінками параметрів. Найкращим буде той метод, для якого вибіркоче середнє помилки оцінювання буде менше.

Отже, МНК найбільш точний для оцінювання параметрів регресійної моделі із шумами, що мають розподіл Гауса, Стюдента з декількома

ступенями свободи, «двогорбий» розподіл на основі гауссівських величин, трикутний розподіл, а також «двогорбий» розподіл на основі трикутного. Цей метод дає найгіршу оцінку при розподілі Лапласа, Коші, Тьюки та Стюдента з менш ніж 4 ступенями свободи.

Література

1. Robustness в статистике. Подход на основе функций влияния/ Хампель Ф., Рончетти Э., Рауссеу П., Штаэль В. -М.: Мир, 1989.
2. Louis A. Jaeckel. Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals [Електронний ресурс]// The Annals of Mathematical Statistics, Vol. 43, Number 5 (1972), pp. 1449-1458, 1972. - Режим доступу http://projecteuclid.org/download/pdf_1/euclid.aoms/1177692377.
3. D. Pollard. Asymptotics for least absolute deviation regression estimators [Електронний ресурс]// Econometric Theory, 7, pp. 186-199, 1991. – Режим доступу:<http://www.math.pku.edu.cn/teachers/xirb/Courses/QR2013/Pollard91ET.pdf>.

Інформаційна технологія формування компетентностей та прогнозування тенденцій вимог до фахівців ІТ галузі

Торчинський О.І., Пасічник О.А

Хмельницький національний університет

Сьогодні інформаційні технології почали активно впливати на повсякденну діяльність будь-якої організації і стали невід'ємною частиною інформаційної інфраструктури цієї організації. Інформаційні технології дозволяють досліджувати та взаємопов'язувати складові ділянки діяльності організації між собою. Вони розвиваються надзвичайно швидкими темпами і захоплюють все ширші сфери діяльності таким чином, що будь-яка конкурентоспроможна діяльність у майбутньому не може бути сформована без детального аналізу можливостей застосування інформаційних технологій. Звідси і випливає, що одним із ключових факторів успішного працевлаштування у сфері інформаційних технологій є володіння певними знаннями та навичками.

В залежності від сфери застосування та специфіки діяльності, кандидату потрібно володіти певним інструментарієм, що є необхідним для виконання майбутніх задач у безпосередній практичній діяльності. Проблема такого роду була, є і буде актуальною завжди в усіх сферах діяльності. Особливо значною вона є для такої динамічної галузі як інформаційні технології. Адже обираючи для вивчення мову програмування чи опановуючи систему керування базами даних потрібно враховувати чи будуть вони затребуваними та актуальними найближчим часом, така сама ситуація з вибором фреймворків, технологій реалізації програмних застосувань, тощо.

На просторах інтернету є різні підходи до вирішення поставленої проблеми: поради від інженерів з власного досвіду, статті написані власниками найбільших ІТ-компаній України, але подібні підходи страждають значним суб'єктивізмом та істотними часовими обмеженнями, оскільки інформація яка була надрукована учора сьогодні вже може бути не актуальною.

Метою роботи є розробка інформаційної технології з формування компетентностей та прогнозування вимог до фахівців ІТ-галузі, що дасть змогу отримувати актуальні дані про перелік затребуваних технологій та їх майбутні перспективи.

Прогнозуванням називається процес передбачення майбутнього стану предмета чи явища на основі аналізу його минулого і сучасного, систематично оцінювальна інформація про якісні й кількісні характеристики розвитку обраного предмета чи явища в перспективі [1].

Широке застосування інформаційних технологій та прискорені темпи їх розвитку викликають необхідність в підготовці висококваліфікованих кадрів, здатних ефективно використовувати нові мови програмування та суміжні з ними технології. Для успішної їх підготовки необхідно знати актуальні на сьогоднішній день та перспективні в майбутньому технології, що використовуються в розробці програмного забезпечення.

Роль прогнозування останнім часом незмірно зростає. Це пов'язано з тим, що в умовах постійного зростання обсягу інформації, збільшення числа факторів, що впливають на суспільний розвиток, необхідно суворе обґрунтування найбільш прогресивних і економічних тенденцій розвитку інформаційних технологій в цілому і його окремих галузей. Прогнозування дозволяє встановити найбільш ймовірні тенденції і напрямки такого розвитку, визначити можливі альтернативи рішення.

Зважаючи на існуючі програмні додатки та медіа-ресурси, була запропонована інформаційна технологія з формування компетентностей та прогнозування вимог до фахівців ІТ галузі.

Вхідними даними для застосування інформаційної технології є мови програмування, які цікавлять кінцевого користувача додатку.

Основу реалізації можна поділити на чотири модулі: parser, cleaner, incrementor та builder рисунок 1.

Перший модуль відповідає за збирання даних. Вхідними даними для даного модуля є введення мови програмування, на основі якої буде відбуватись пошук суміжних технологій. Збирання даних відбувається з таких сайтів, як «www.work.ua» та «www.rabota.ua».

Другий модуль виконує роль очисника, який аналізує кожен зібрану вакансію, та відділяє непотрібні слова, від потенційних технологій, що також зменшує розмір майбутнього тексту для аналізу. Так як нові технології з'являються дуже активно, було обрано підхід з відкиданням слів, які не є

предметно-орієнтованими, а не звичайне підрахування з уже існуючих технологій.

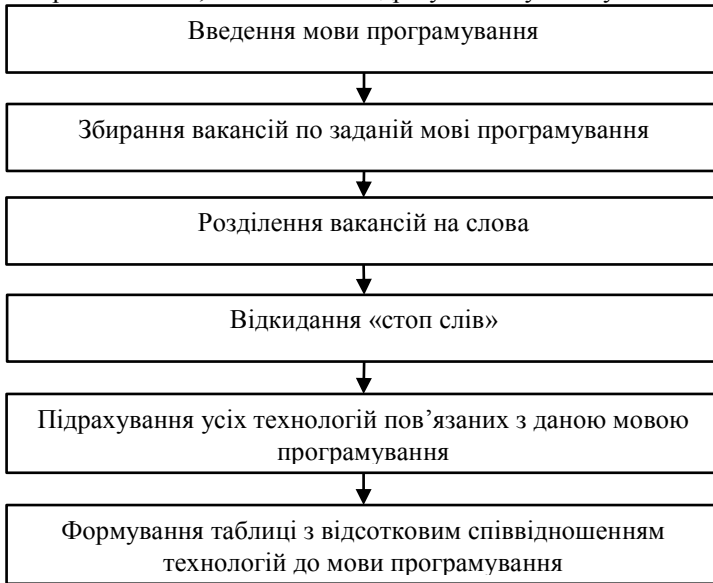


Рисунок 1 – Схема роботи інформаційної технології формування компетентностей та прогнозування тенденцій вимог до фахівців ІТ галузі

Третій модуль займається обробкою кожної зібраної вакансії, після стадії очистки в модулі cleaner. На даному етапі ми маємо лише мову програмування по якій виконувався пошук, та список усіх суміжних з нею технологій. Кожна технологія, має певну кількість відносно загальної кількості вакансій, на основі цієї кількості можна зробити припущення, щодо її значущості поряд з даною мовою програмування.

Четвертий являє собою будівник – це модуль, який виконує функцію агрегатора, відображає мову програмування по якій проводиться пошук, та перелік суміжних з нею технологій з пороговим значенням входження 5 %.

У якості сховища даних було використано NoSql базу даних MongoDB. MongoDB являє собою документ-орієнтовану систему керування базами даних, яка не потребує опису схеми таблиць.

Переваги MongoDB:

Динамічна схема: Як згадувалося вище, ця СКБД дозволяє гнучко працювати зі схемою даних без необхідності змінювати самі дані;

Масштабованість: MongoDB горизонтально масштабована, що дозволяє легко зменшити навантаження на сервера при великих обсягах даних;

Швидкість: Висока продуктивність при виконанні простих запитів;

Гнучкість: В MongoDB можна без шкоди для існуючих даних, їх

структури і продуктивності СКБД додавати поля або колонки.

Її UML представлення має вигляд, що зображено на рисунку 2.

Отже, запропонована інформаційна технологія формування компетентностей та прогнозування тенденцій вимог до фахівців ІТ-галузі забезпечує можливість отримання найбільш необхідних та актуальних навичок та знань потрібних для розробника певного профілю.

Подальші дослідження спрямовані на спостереження та аналіз залежностей між технологіями та виявлення найбільш швидко зростаючих, а також розширення джерел отримання інформації для аналізу технологій.

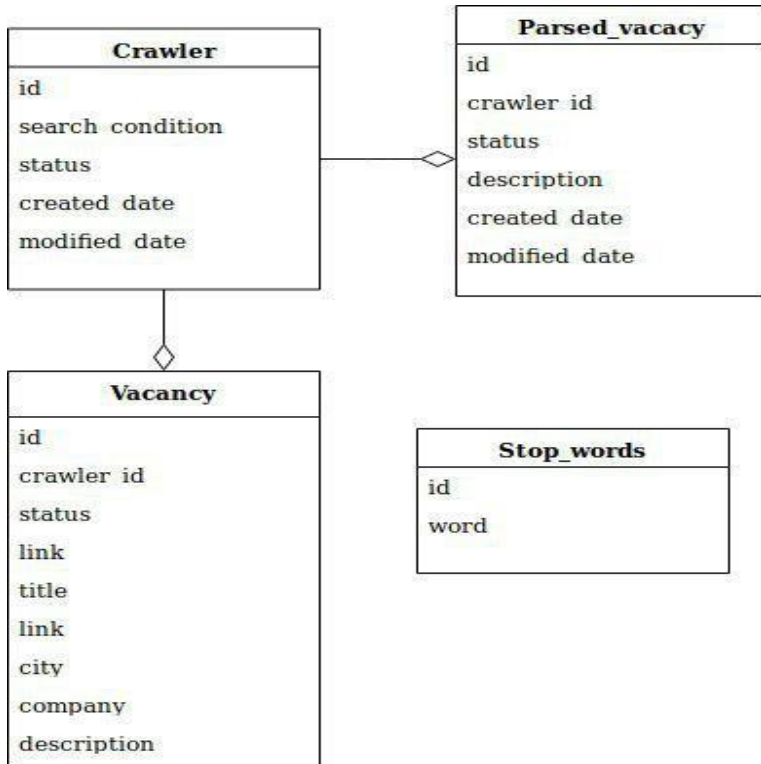


Рисунок 2 - Представлення схеми бази даних

Література

1. Технічна енциклопедія TechTrend– Open-source learning platform. – [Електронний ресурс]. – Режим доступу: techtrend.com.ua/index.php?newsid=26003804
2. Особливості застосування нових інформаційних технологій в управлінні персоналом. [Електронний ресурс]. – Режим доступу: <http://oldconf.neasmo.org.ua/node/720>

Аналіз методу тадеуса вінсенті для вирішення прямої геодезичної задачі на еліпсоїді

Фарина А.П.

Науковий керівник – д.т.н., проф. Мясіщев О.А.

Хмельницький національний університет

Часто в навігації чи при проектуванні інженерних споруд, іноді і в інших галузях, виникає потреба розрахувати координати точки. Цю проблему може вирішити пряма геодезична задача.

Прямою задачею в геодезії називають задачу, в якій по даним координатам однієї точки, дирекціонному куту (азимуту) напрямку з цієї точки на іншу, та відстані між ними, знаходять координати іншої точки та дирекційний кут (азимут) зворотного напрямку.

Одним з варіантів практичного застосування – є розрахунок апріорної GPS координати в навігаційній системі оснований на фільтрі Калмана.

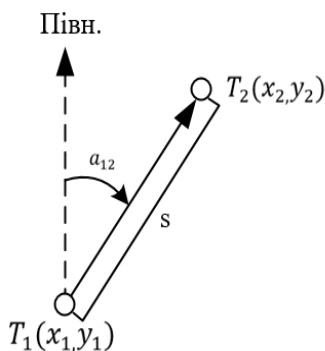


Рисунок 1 – пряма геодезична задача

Класичне рішення задачі:

1. Знаходимо координати невідомої точки:

$$x_2 = x_1 + s \cdot \cos a_{12}$$

$$y_2 = y_1 + s \cdot \sin a_{12}$$

2. Визначаємо значення кута зворотного напрямку (азимуту):

$$a_{21} = \left| a_{12} + 180^\circ \right|_{360^\circ}$$

Однак ці формули дозволяють визначити тільки приблизне значення координати з великою похибкою.

Формули Тадеуса Вінсенті дозволяють визначити координати точки з сантиметровою точністю, але вимагають більш складних розрахунків. Метою Вінсенті було вираження існуючих алгоритмів геодезичних на еліпсоїді у короткій формі для швидкого обчислення і малого використання пам'яті. Для отримання високої точності використовується класичне рішення Лежандра (1806), Бесселя (1825) та Гельмерта (1880) на основі допоміжної сфери. (Вінсенті спирався на формулювання методу, даного Ренсфордом, 1955 р.). Лежандр показав, що еліпсоїдальна геодезична форма може бути точно віднесена до великого кола на допоміжну сферу шляхом відображення географічної широти до зменшеної широти і встановлення азимута великого кола рівним до геодезичного.

Формули методу:

$$a = 6378137; b = 6356752.314245; f = (a - b) / a;$$

$$\tan U = (1 - f) \tan \varphi_1; \cos U = 1 / \sqrt{1 + \tan^2 U}; \sin U = \tan U \cdot \cos U;$$

$$\sigma_1 = a \tan(\tan U / \cos a_1); \sin a = \cos U \cdot \sin a_1; \cos^2 a = 1 - \sin^2 a;$$

$$u^2 = \cos^2 a \cdot (a^2 - b^2) / b^2;$$

$$A = 1 + u^2 / 16384 \cdot (4096 + u^2 \cdot (-768 + u^2 \cdot (320 - 175 \cdot u^2)));$$

$$B = u^2 / 1024 \cdot (256 + u^2 \cdot (-128 + u^2 \cdot (74 - 47 \cdot u^2)));$$

do {

do {

$$\cos 2\sigma_m = \cos(2\sigma_1 + \sigma);$$

$$\sigma = B \cdot \sin \sigma \cdot \left(\cos 2\sigma_m B / 4 \cdot \left(\cos \sigma \cdot (-1 + 2 \cdot \cos^2 2q_m) - B / 6 \cdot \cos 2\sigma_m \right) \right. \\ \left. \cdot (-3 + 4 \cdot \sin^2 \sigma) \cdot (-3 + 4 \cdot \cos^2 2q_m) \right)$$

$$\sigma = s / b \cdot A + \sigma;$$

$$\} \text{WHILE } |\sigma - s / (b \cdot A)| < 1 \cdot 10^{-12};$$

$$\varphi_2 = a \tan(\sin U \cdot \cos \sigma + \cos U \cdot \sin \sigma \cdot \cos a_1) / (1 - f) \cdot$$

$$\sqrt{\sin^2 a + (\sin U \cdot \sin \sigma - \cos U \cdot \cos \sigma \cdot \cos a_1)^2};$$

$$\lambda = a \tan(\sin \sigma \cdot \sin a_1 / \cos U \cdot \cos \sigma - \sin U \cdot \sin \sigma \cdot \cos a_1);$$

$$C = f / 16 \cdot \cos^2 a \cdot (4 + f \cdot (4 - 3 \cdot \cos^2 a));$$

$$\lambda_2 = \lambda_1 + \left(\lambda - (1 - C) \cdot f \cdot \sin a \cdot \left(\frac{\sigma + C \cdot \sin \sigma \cdot (\cos 2\sigma_m + C \cdot \sin \sigma \cdot (-1 + 2 \cdot \cos^2 2q_m))}{C \cdot \sin \sigma \cdot (-1 + 2 \cdot \cos^2 2q_m)} \right) \right)$$

$a_2 = \arctan(\sin a / -\sin U \cdot \sin \sigma + \cos U \cdot \cos \sigma \cdot \cos a_1)$, де a – велика вісь еліпсоїда землі, b – мала вісь еліпсоїда землі, U – зменшена широта, σ – довжина дуги між точками на допоміжній сфері, φ – широта, λ – довгота, a_1 і a_2 – азимуту.

Література

1. Геодезичні задачі: пряма та обернена задачі. [Electronic resource]. – Mode of access: <http://www.geoguide.com.ua/survey/survey.php?part=geod&art=geod010>
2. Формули Вінценті. [Electronic resource]. – Mode of access: https://en.wikipedia.org/wiki/Vincenty%27s_formulae
3. Вирішення геодезичних задач на еліпсоїді. [Electronic resource]. – Mode of access: <http://www.movable-type.co.uk/scripts/latlong-vincenty.html>

Інформаційна технологія для збільшення кількості переглядів оголошень з нерухомості

Шкута В.О., Багрій Р.О., Скрипник Т.К.
Хмельницький національний університет

З розвитком інформаційних технологій процес публікації оголошень перемістився з газет та дошок на електронні ресурси. Основна причина зацікавленості аудиторії в електронних системах пошуку оголошень – це економія часу та зусиль на пошук потрібної послуги чи товару. Однією з найпопулярніших категорій оголошень в Україні є нерухомість.

Оголошенням називають коротке текстове повідомлення, яке містить різну за призначенням інформацію. Переважно оголошення містять інформацію приватного рекламного характеру – це пропозиція певних товарів чи послуг приватними особами або малими приватними підприємствами. Комерційні оголошення є одним із видів доступної та дешевої реклами, тому вони дуже широко використовуються в мережі Internet, на віртуальних дошках оголошень [1].

При написанні оголошення з нерухомості користувачі часто зустрічаються з проблемою недостатньої популярності оголошення. Це, в свою чергу, негативно відображається на реалізації поставленої мети. Для того, щоб оголошення зацікавило потенційного клієнта, потрібно дотримуватись певних вимог.

Метою роботи є розробка інформаційної технології, яка б дала можливість підвищити кількість переглядів оголошень, зробити їх більш інформативними та компактними. Предмет дослідження – оголошення з нерухомості, що розміщуються в мережі Internet.

Наукові дослідження стверджують, що на кількість переглядів статей впливає психологічний фактор. Так, в роботі [2], було проведено аналіз тональності заголовків (69907 заголовків статей чотирьох міжнародних засобів масової інформації за 2014 рік) на предмет позитивних, негативних і нейтральних емоцій. За результатами дослідження було виявлено, що вкрай негативні або вкрай позитивні новини, як правило, залучають більшу кількість читачів. Більше того, заголовок має більше шансів отримати перегляди, якщо почуття, що виражаються в тексті, є надлишковими, при цьому, не важливо чи вони позитивні чи негативні.

Для вирішення поставленої мети, була запропонована інформаційна технологія. В основі функціонування інформаційної технології є використання набору правил для створення повноцінного відформатованого оголошення, з вхідного тексту оголошення. Схема роботи даної технології зображена на рисунку 1.

На першому етапі з тексту оголошення визначаються ключові слова, потім додаються клікбейт слова [3], визначаються атрибути об'єкту

нерухомості, на останньому етапі, з отриманих даних формується заголовок оголошення та відформатований текст оголошення.

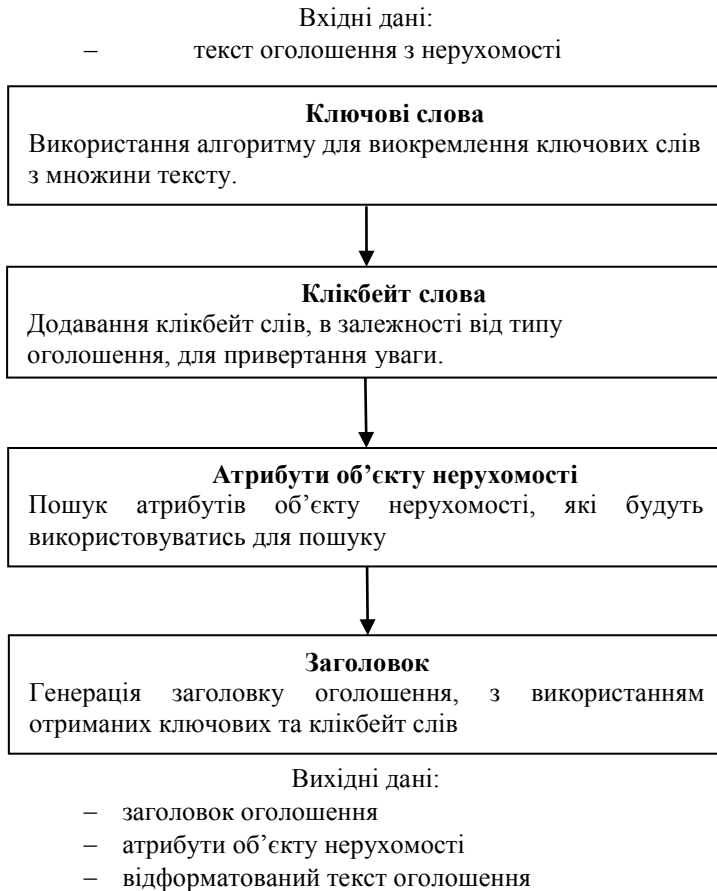


Рисунок 1 – Схема інформаційної технології для підвищення переглядів оголошень з нерухомості

Ключовим словом називають слово в тексті, яке використовують для вираження деякого змісту документу та яке, в сукупності з іншими ключовими словами, здатне описати текст документу. Набір ключових слів близький до анотації, плану та конспекту, які теж описують документ з меншою деталізацією, але він позбавлений синтаксичної структури. Проблема пошуку ключових слів є однією з найважчих задач лінгвістики

тексту, тому, на даний момент, існує велика кількість алгоритмів автоматичного пошуку ключових слів з тексту на різних мовах.

Майже всі сучасні алгоритми для пошуку ключових слів використовують видалення стоп-слів в процесі обробки тексту. Стоп-словами називають слова, які не несуть сенсового навантаження, тому їх користь та роль для пошуку не суттєва. Стоп-слова діляться на загальні та залежні. До загальних можна віднести прийменники, суфікси, дієприкметники, вигуки, цифри тощо. Залежними називають ті слова, залежать від пошукової фрази.

Одними з найпопулярніших алгоритмів для пошуку ключових слів є граф-орієнтовані алгоритми «TextRank» та «Rake» [4]. Дані алгоритми представляють текст як безліч слів-вершин та ребер, що є відносинами між цими словами.

Отже, на першому етапі відбувається виокремлення ключових слів з множини тексту, які в подальшому необхідні для генерації назви оголошення.

Наступним етапом є додавання клікбейт слів. Клікбейт – це спосіб залучення аудиторії за допомогою специфічних заголовків, які в певних випадках супроводжуються графічними матеріалами, що провокують інтернет-користувачів переглядати певний контент. Явище клікбейту має дещо негативний характер тому, що часто за допомогою нього користувачів вводять в оману. Основне завдання клікбейту - викликати як негативні, так і позитивні емоції, для того щоб привернути увагу. Як відомо, саме емоційне збудження є одним з ключових факторів, що визначають поведінку людини в тій чи іншій ситуації.

Третім етапом є виділення атрибутів об'єкту нерухомості з оголошення. Атрибутами оголошення називають основні характеристики предмету оголошення. У випадку оголошень з нерухомості це характеристика об'єкту (ціна, кількість кімнат, і т.д.). Ці дані потрібні для точнішого та швидкого пошуку.

З отриманих даних на попередніх етапах формується заголовок оголошення. Зазвичай, в системах пошуку оголошень заголовком це те, що користувач бачить першим, детальний опис прихований до тих пір, поки не відбудеться перехід по назві. Заголовок оголошення має містити коротку, інформативну та привабливу назву товару або послуги. Згідно з теорією «Магічне число сім, плюс-мінус два» [5], середньостатистична людина здатна тримати в робочій пам'яті сім об'єктів. Саме тому, оптимальною довжиною заголовку оголошення вважається сім або менше слів.

Приклад формування заголовку з тексту оголошення показано на рисунку 2.

Після виконання всіх етапів інформаційної технології формується оголошення, яке включає в себе заголовок оголошення, атрибути об'єкту нерухомості та відформатований текст оголошення.

Текст оголошення

Продам 3 кімнатну квартиру, з євроремонтом, в центрі Києва. Квартира здається вперше. Є просторий балкон, присутні лічильники. Будинок має дуже зручне розташування - поряд, банки, школи, дитячі садочки. Можливий продаж з меблями та без. Ціна 80 тисяч \$

Алгоритм генерації заголовку

- 1. Пошук ключових слів:** продаж, квартиру, з євроремонтом, вперше, балкон, лічильники, будинок, зручне, банки, школи, дитячі, меблями, ціна
- 2. Додавання клікбейт слів:** терміново, затишна
- 3. Атрибути оголошення:** рубрика – продаж, об'єкт – квартира, місто – Київ, ціна – 80.000 \$, додаткові опції – балкон, лічильники, меблі

Заголовок оголошення

Терміново! Продам затишну 3 кімнатну квартиру.

Рисунок 2 – Приклад генерації заголовку оголошення

Отже, запропонована інформаційна технологія для підвищення переглядів оголошень з нерухомості забезпечує вдосконалення та приведення оголошення до оптимального формату, завдяки чому воно стає більш популярним, що збільшує ймовірність в досягненні поставленої мети.

Подальші дослідження спрямовані на пошук альтернативних методів підвищення переглядів, збільшення переліку клікбейт слів та вдосконалення методики генерації заголовків оголошення з нерухомості, що в свою чергу позитивно відобразиться на кількості переглядів.

Література

1. Оголошення – матеріал з Вікіпедії [Електронний ресурс]. – Режим доступу: <https://uk.wikipedia.org/wiki/Оголошення>
2. Julio Reis, Fabricio Benevenuto, Pedro O.S. Vaz de Melo, Raquel Prates, Haewoon Kwak, Jisun An. Breaking the News: First Impressions Matter on Online News [Електронний ресурс]. – Режим доступу: <https://arxiv.org/pdf/1503.07921v2.pdf>
3. Клікбейт – матеріал з Вікіпедії [Електронний ресурс]. – Режим доступу: <https://uk.wikipedia.org/wiki/Клікбейт>
4. Stuart Rose, Dave Engel, Nick Cramer, and Wendy, Cowley 2010. Automatic keyword extraction from individual documents. C.3-20.
5. The Magical Number Seven [Електронний ресурс] – Режим доступу: http://www.psych.utoronto.ca/users/peterson/psy430s2001/Miller_GA_Magical_Seven_Psych_Review_1955.pdf

Особливості використання методу випадкового лісу для визначення шаблонів в потоці даних

Шлапак О.В., Манзюк Е.А.

Хмельницький національний університет

Random Forest (RF) із застосуванням беггінгу – один з найкращих методів машинного навчання, який небагато уступає градієнтному бустингу.

Випадковий ліс складається з комітету дерев розв'язків (які також називаються деревами класифікації або регресійними деревами "CART" і вирішують однойменні завдання). Вони застосовуються в статистиці, аналізі даних і машинному навчанні. Кожне окреме дерево – достатньо проста модель, яка має вузли та листки. У вузлах записані атрибути, від значень яких залежить цільова функція. Далі по вітках у листки попадають значення цільової функції [1-3]. У процесі класифікації нового випадку потрібно спуститися по дереву через вітки до листків, пройшовши через усі значення атрибутів по логічному принципу "ЯКЩО-ТО". Залежно від цих умов, цільовій змінній буде привласнено те або інше значення або клас (цільова змінна потрапить у конкретну вітку). Ціль побудови дерева розв'язків – створення моделі, яка прогнозує значення цільової змінної залежно від декількох змінних на вході.

Випадковий ліс будується шляхом простого голосування дерев розв'язків по алгоритму Bagging (Беггінг). Bagging – це штучне слово, утворене від англійського словосполучення «bootstrap aggregating». Цей термін увів Лео Брейман в 1994 році.

Bootstrap в статистиці – це такий спосіб формування вибірки, коли вибирається рівно стільки ж об'єктів, скільки їх на початку та було. Але об'єкти ці вибираються з повтореннями. Іншими словами, обраний випадковий об'єкт вертається назад і може бути обраний повторно. При цьому число об'єктів, яке буде обрано, складе приблизно 63% від вихідної вибірки, а частка, що залишилася, об'єктів (приблизно 37%) жодного разу не потрапить у навчальну вибірку. По цієї сформованої семплованої вибірці навчаються базові алгоритми (у нашому випадку дерева розв'язків). Це теж відбувається випадковим чином: беруться випадкові підмножини (семпли) заданої довжини та навчаються на обраній випадковій підмножині ознак (атрибутів). А 37% вибірки, що залишилися, використовуються для перевірки узагальнюючої здатності побудованої моделі.

Потім усі навчені дерева поєднуються в композицію за допомогою простого голосування з використанням усередненої помилки для всіх семплів. У результаті застосування bootstrap aggregating зменшується середньо квадратична помилка, знижується дисперсія класифікатора, якого навчають. На різних вибірках помилка буде відрізнятися не так сильно. У результаті модель буде менше перенавчатися. Ефективність беггінгу полягає в тому, що базові алгоритми (дерева розв'язків) навчаються по різним

випадковим підвибірках і їх результати можуть сильно відрізнятись, але їх помилки взаємно компенсуються при голосуванні.

Метою роботи є дослідження можливості застосування випадкового лісу для апроксимації якої-небудь функції. У даному випадку функцією від показань індикаторів був сигнал на здійснення дій. Також було розглянуто два види навчання із учителем – класифікація та регресія, і застосований нестандартний підхід до навчання із заздалегідь невідомими навчальними прикладами, які підбиралися автоматично, у процесі оптимізації. У цілому, процес схожий на звичайну оптимізацію в тестері, при цьому алгоритм сходиться всього за кілька ітерацій і не варто піклуватись про кількість параметрів оптимізації. Це очевидна перевага застосованого підходу. Мінус моделі – сильний оверфіт при неправильно обраній стратегії, який еквівалентний переоптимізації параметрів.

Загалом random forest – це спеціальний випадок бегінгу, коли в якості базового сімейства використовуються вирішальні дерева. При цьому, на відміну від звичайного способу побудови дерев рішень, не використовується усікання дерева (pruning). Метод налаштований на те, щоб можна було побудувати композицію якнайшвидше по великих вибірках даних. Кожне дерево будується специфічним чином. Ознака (атрибут) для побудови вузла дерева вибирається не із загального числа ознак, а з їхньої випадкової підмножини. Якщо будуємо регресійну модель, то число ознак рівно $n/3$. У випадку класифікації це \sqrt{n} . Усе це є емпіричними рекомендаціями та називається декореляцією: у різні дерева попадають різні набори ознак, і дерева навчаються на різних вибірках.

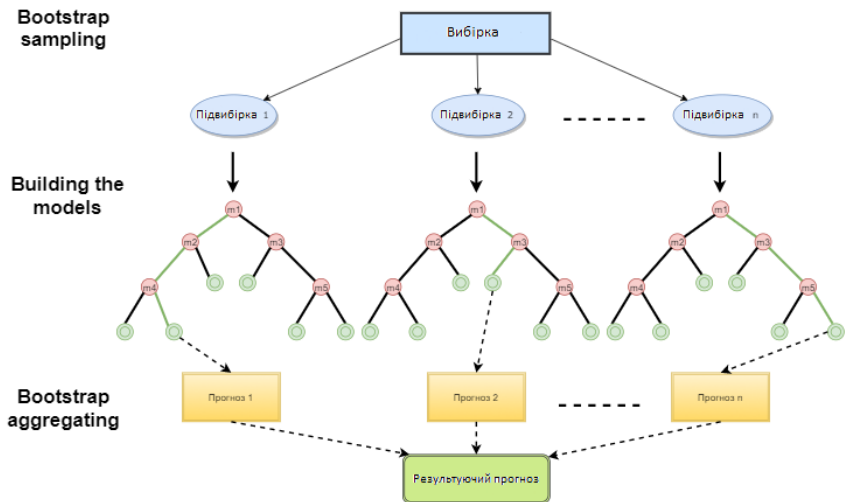


Рисунок 1 - Схема функціонування методу випадкового лісу

Алгоритм випадкового лісу виявився надзвичайно ефективним, здатним вирішувати практичні завдання. Він дає високу точність навчання при великій кількості випадковостей, внесених у процес побудови моделі. Перевага перед іншими моделями машинного навчання – оцінка out of bag для частини множини, що не потрапила в навчальну вибірку. Тому для дерев розв'язків не обов'язково проводити крос-валідацію або тестування на окремій вибірці. Досить обмежитися оцінкою out of bag для подальшого налаштування моделі: добору кількості вирішальних дерев і регуляризуючої складової.

Переваги алгоритму:

- висока швидкість навчання;
- неітеративне навчання – алгоритм завершується за фіксоване число операцій;
- масштабованість (здатність обробляти великі обсяги даних);
- висока якість одержуваних моделей (порівнюється з нейронними мережами та ансамблями нейронних мереж);
- відсутність чутливості до викидів у даних через випадкове семплування;
- мала кількість параметрів, що настраюються;
- відсутність чутливості до масштабування (і взагалі до будь-яких монотонних перетворень) значень ознак, завдяки вибору випадкових підпросторів;
- не вимагає ретельного налаштування параметрів. За допомогою налаштування параметрів можна досягти приросту від 0.5 до 3% точності залежно від завдання та даних;
- добре працює із пропущеними даними – зберігає гарну точність, навіть якщо більша частина даних пропущена;
- внутрішня оцінка здатності моделі до узагальнення;
- можливість роботи із сирими даними, без препроцесінгу.

Перша та основна проблема – вибір того, чому конкретно потрібно навчати модель. На ціну впливає безліч факторів. Є тільки один науковий підхід вивчення ринку – економетрика. У ній три основні методи: регресійний аналіз, аналіз тимчасових рядів і панельний аналіз. Окремий цікавий розділ цієї науки – непараметрична економетрика, яка ґрунтується винятково на наявних даних, без аналізу причин, їх формуючих. Методи непараметричної економетрики останнім часом популярні в прикладних дослідженнях: приміром, це ядерні методи та нейронні мережі. До цього ж розділу відноситься математичний аналіз нечислових понять – наприклад, нечіткі множини. У цьому дослідженні зупинимося на нечітких множинах.

Model selection (вибір моделі).

Друга проблема – вибір моделі навчання. Існує безліч лінійних і нелінійних моделей. З їхнім списком і порівняльними характеристиками

можете ознайомитися, наприклад, на сайті Microsoft.

Приміром, є величезна множина різновидів нейронних мереж. Використовуючи їх, доводиться експериментувати з архітектурою мережі, підбираючи кількість шарів і нейронів. У той же час класичні нейромережі типу MLP навчаються повільно (особливо на градієнтному спуску з фіксованим кроком), експерименти з ними займають багато часу. Сучасні швидкі Deep Learning мережі на даний момент не представлені в стандартній бібліотеці терміналу, а підключати сторонні бібліотеки у вигляді dll не зовсім зручно.

На відміну від нейронних мереж, лінійні моделі працюють швидко, але не завжди добре справляються з апроксимацією.

Random Forest позбавлений цих недоліків. Таким чином потрібно лише підібрати кількість дерев і параметр g , відповідальний за відсоток об'єктів, що потрапили в навчальну вибірку. Випадкові ліси в більшості випадків дають найменшу помилку прогнозування на тестових даних.

Література

1. Evans D. A., Lefferts R. G. Clarit-trec experiments // Information processing & management. 1995. Vol. 31(3). P. 385–395.
2. Frantzi K., Ananiadou S., Mima H. Automatic recognition of multi-word terms: the c-value/nc-value method // International Journal on Digital Libraries. 2000. Vol. 3(2). P. 115–130.
3. Govindaraju V., Ramanathan. K. Similar Document Search and Recommendation // Journal of Emerging Technologies in Web Intelligence. 2012. Vol. 4(1). P. 84–93.

Автоматизація побудови гібридної інфраструктури на базі Microsoft Exchange Server та Office 365

Шоханов А.С., Міхалевський В.Ц.

Хмельницький національний університет

Популяризація ІТ-інфраструктури на базі гібридних хмар надає нових можливостей не лише споживачам, а й провайдерам послуг. Вони теж отримують переваги, адже зможуть надавати своїм клієнтам гнучке та надійне середовище, яке задовольняє їхні вимоги, з високим рівнем безпеки включно.

Хмарне середовище значно змінює спосіб роботи бізнесу, проте це не зменшує вимог до захищеності. У результаті безпека стала основною і потенційно найприбутковішою для хостерів, які працюють з «хмарами». Хостинг де-факто став рішенням для «довіреної хмари», і споживачі готові платити додатково за гарантії.

Перехід до хмарної інфраструктури може бути складним і тому Microsoft допомагає споживачам його здійснити. Цьому сприяє широке

портфолію відповідних продуктів корпорації: серверні рішення Windows Server, Microsoft System Center, Microsoft SQL Server, Windows Azure, а також програми Exchange, SharePoint, Skype for Business, Office та Office 365. Із системою Cloud OS Microsoft пропонує бізнесу комплексне рішення як для побудови хмарної інфраструктури, так і впровадження складних гібридних інфраструктур для об'єднання локальної та хмарної інфраструктури. Корпорація також надає споживачам максимальну гнучкість у використанні цих технологій.

Одним з найважливіших сервісів, яке складає основу будь-якого підприємства є рішення Exchange Server, яке допомагає забезпечити постійну доступність і керуваність зв'язку. Корпоративна електронна пошта, календарі та контакти на ПК, телефоні і в браузері - завдяки спрощеному підходу до підвищення доступності та аварійного відновлення ви досягаєте нових рівнів надійності з меншими зусиллями і забезпечуєте захист конфіденційної інформації. Основні функції Microsoft Exchange: обробка і пересилка поштових повідомлень, спільний доступ до календарів і задач, підтримка мобільних пристроїв, захист інформації від крадіжки і веб-доступ, інтеграція із системами голосових повідомлень, підтримка систем обміну миттєвими повідомленнями.

При використанні рішень Office 365 компанії отримують доступ по знайомих засобах Office практично звідусіль. Microsoft Office 365 пропонує сучасні технології для спільної роботи та об'єднаних комунікацій організаціям будь-якого розміру. Завдяки вдалому поєднанню функціонала Microsoft Office з онлайновими сервісами Microsoft новий пакет Office 365 допомагає клієнтам заощадити гроші і час, а також оптимізувати навантаження на ІТ-персонал, адже сервіси Office 365 зручні в користуванні і адмініструванні і мають високий рівень безпеки та надійності.

Microsoft Office 365 – це платний хмарний власницький інтернет-сервіс і програмне забезпечення компанії Microsoft, що розповсюджується за схемою «програмне забезпечення + послуги» (англ. Software + Services). Хмарний формат означає, що дані зберігаються в центрі обробки даних, а не на комп'ютері, що забезпечує користувачам доступ до документів і даних через браузер з різних пристроїв з можливістю виходу в Інтернет.

Сервіс Office 365 був анонсований в жовтні 2010 року, публічне бета-тестування розпочалося у квітні 2011. Світова прем'єра відбулася 28 червня 2011 року, коли корпорація Microsoft запустила сервіс Office 365 у 40 країнах світу. Пакет призначений для використання в компаніях – від малого бізнесу до великих підприємств. Середньому та малому бізнесу він буде надаватися через партнерів корпорації, а великими клієнтами Microsoft займається власноруч в рамках програми корпоративного ліцензування Enterprise Agreement, яка включає ліцензування базових продуктів Microsoft для всіх використовуваних ПК (від 250 робочих місць).

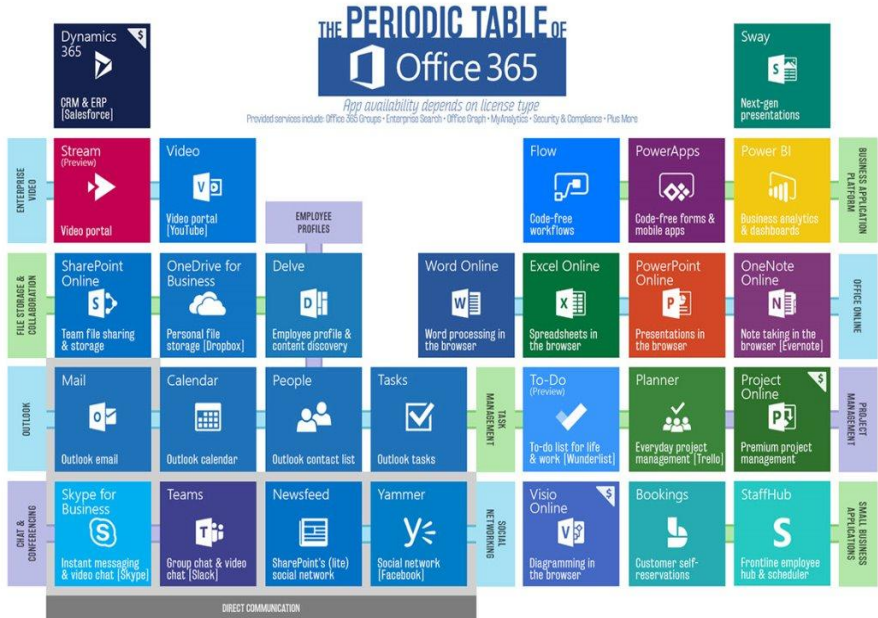


Рисунок 1 – Таблиця сервісів хмарної служби Office 365

Як було зазначено вище, хмарний офісний пакет послуг Microsoft Office 365 включає в себе такі найважливіші продукти: 1) Microsoft Office Professional Plus, що забезпечує можливість роботи з документами в знайомому інтерфейсі застосувань Office на комп'ютері, телефоні або через веб-браузер; 2) Exchange Online, що дозволяє розгорнути у хмарі сервіси електронної пошти Outlook, календаря і контактів і забезпечує захист від вірусів і спаму; 3) SharePoint Online для створення веб-порталів організації і внутрішніх соціальних мереж для спілкування та взаємодії співробітників; 4) Skype for Business Online для організації відео- і голосових конференцій з колегами і партнерами, а також можливість налаштування та використання програм обміну миттєвими повідомленнями. Але останнім часом Microsoft пропонує новий сервіс Teams, який інтегрував в собі функції конференцій, чату з архівом, сховища файлів, та багато інших надбудов, для того, щоб організувати колективну роботу в рамках проекту.

Сервіс може інтегруватися з корпоративними ERP і CRM-системами або надавати доступ к хмарним аналогам.

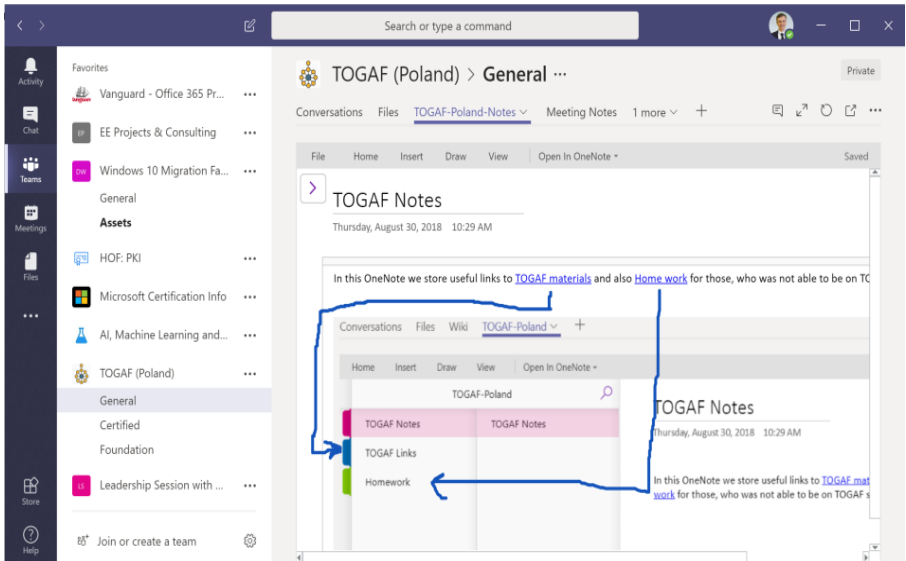


Рисунок 2 – Приклад роботи в Microsoft Teams

Практична реалізація побудови гібридної інфраструктури між наземною та хмарною поштовою організаціями полягає в реалізації багатьох кроків: 1) Інвентаризація інфраструктури мережевої та організаційної; 2) Аналіз мережевої інфраструктури та Активного Каталогу; 3) Аналіз транспортної системи; 4) Аналіз даних користувачів поштової системи; 5) Аналіз клієнтського доступу до системи; 6) Аналіз елементів безпеки; 7) Аналіз інших елементів; 8) Формування звіту про знайдені елементи інфраструктури поштової організації; 9) Корегування архітектури під специфічні умови підприємства; 10) Аспекти, що впливають на транспортні елементи; 11) Аспекти, безпеки, які впливають на дизайн і майбутню конфігурацію гібридної інфраструктури; 12) Побудова інфраструктури; 13) Підготовка глобальної конфігурації Організації Exchange, для інтеграції з хмарним сервісом Office 365; 14) Впровадження гібридної інфраструктури; 15) Перенос елементів конфігурації в хмарну частину гібридної інфраструктури; 16) Налаштування елементів безпеки в хмарному середовищі; 17) Автоматичне тестування та формування звіту про впровадження гібридної інфраструктури; 18) Створення тестових ресурсів інфраструктури для тестування; 19) Тестування здоров'я існуючої інфраструктури; 20) Тестування міграції даних; 21) Тестування клієнтської частини; 22) Тестування транспорт-ного функціоналу.

В результаті, побудова гібридної інфраструктури на базі Microsoft Exchange Server та Office 365 надає суттєві переваги. При використанні

хмарних обчислень, споживачі інформаційних технологій можуть істотно знизити капітальні витрати – на побудову центрів обробки даних, закупівлю серверного та мережевого обладнання, апаратних і програмних рішень щодо забезпечення безперервності і працездатності –так як ці витрати поглинаються провайдером хмарних послуг. Крім того, тривалий час побудови та введення в експлуатацію великих об'єктів інфраструктури інформаційних технологій та висока їх початкова вартість обмежують можливість гнучкого реагування на потреби ринку, тоді як хмарні технології забезпечують можливість практично миттєво реагувати на збільшення попиту на обчислювальні потужності. При використанні хмарних обчислень, витрати споживача зміщуються в бік операційних –таким чином компенсуються витрати на оплату послуг хмарних провайдерів.

На сьогоднішній день процес міграції даних займає дуже багато часу, так як вимагає ретельного аналізу внутрішньої інфраструктури, підготовки її для міграції, побудову гібридної інфраструктури, де частина призначених для користувача даних може знаходитися в хмарі, а частина в наземному ЦОДі і безпосередньо перенесення даних. Щоб вирішити ці проблеми компанії звертаються за допомогою ІТ компаній, які спеціалізуються на перенесенні даних з внутрішнього ЦОДа в хмарний сервіс. Подібні проекти можуть займати дуже багато часу (місяці і навіть роки), із залученням висококваліфікованих фахівців в різних областях, таких як мережева інфраструктура, поштові служби, WEB служби, безпека та ін. Все це в кінцевому підсумку виливається в багато мільйонні проекти, завдання яких лише перенести дані і конфігурацію з однієї локації в іншу. Всі ці дії носять закономірний та логічний характер, що дозволяє зробити автоматизовану систему на базі існуючого API, який планується використовувати в даній роботі для автоматизації всіх 25 вищезазначених кроків та корегування сценарію міграції під потреби компанії.

Література

1. Інформаційна інфраструктура – Вікіпедія [Електронний ресурс] // Вікіпедія. – Режим доступу : <https://goo.gl/9L7xCN>.
2. Модло Є. О. Використання десктопних програм у хмарному середовищі / Є. О. Модло // Хмарні технології в освіті: матеріали Всеукр. Наук.-метод. Інтернет-семінару (Кривий Ріг – Київ – Черкаси – Харків, 21 грудня 2012 р.). – Кривий Ріг: Видавничий відділ КМІ, 2012. – С. 39.
3. Риз Дж. Облачные вычисления / Джордж Риз. – СПб.: БХВПетербург, 2011. – 288 с

Інформаційна технологія для формування автоматизованих корпусів з мережі Інтернет

Яцунь В.В.

Хмельницький національний університет

За останні кілька років в області інформаційного веб-пошуку все частіше проводяться дослідження [1, 4], пов'язані з науковим напрямком, що розбудовується, вебметрика (webometrics). Зокрема, методами вебметрики досліджуються такі питання, як оцінка присутності інформаційних веб-ресурсів у мережі, ранжування веб-сайтів університетів і наукових установ [3] або підвищення вебметричного рейтингу, введеного іспанською групою Cybermetrics Lab (<http://www.webometrics.info>), різних університетів миру.

До актуальних напрямків вебметрики відносяться дослідження з аналізу та виявленню гіперпосилальних структур різних сегментів веб-простору (наприклад, академічний сегмент веба, корпоративний і ін.). Для одержання великих обсягів інформації про гіперпосилання використовуються веб-краулери. Спільним завданням таких інструментів є спеціалізований обхід веба з метою збору інформації або визначення гіперпосилальної структури яких-небудь інформаційних ресурсів [1, 2]. Наявні в наш час веб-краулери можна розділити на дві групи: універсальні та спеціалізовані (тематико-орієнтовані). Універсальні зазвичай є надлишкові, сильно навантажують мережеві ресурси та програють спеціалізованим за такими показниками, як час обходу веб-простору, продуктивність обробки інформації, а також можливість спрямованого пошуку інформаційних джерел у рамках певної метрики значимості. У цьому змісті переваги спеціалізованих систем очевидні, однак їх мінусом є значні витрати на виготовлення - для кожної спеціалізації потрібний свій краулер.

Разом з тим, значне число функцій усіх краулерів збігається [1-4].

У даній роботі пропонується модель спеціалізованих веб-краулерів спеціалізованих доповнень, що полягають із узагальненого ядра та дозволяють сформувати корпус. При такому підході витрати на виготовлення спеціалізованого веб-краулера мінімізуються. Пропонується архітектура узагальненого ядра пошукового робота і її реалізація. Ставиться експеримент, у якому рівняється якість роботи веб-краулерів на основі узагальненого ядра з якістю роботи закордонних аналогів на прикладі Heritrix, Openwebspider і Methanol Web Crawler.

Архітектура узагальненого ядра. Масштабність веб-простору породжує ряд проблем, що впливають на ефективність роботи веб-краулерів будь-яких типів. Аналіз відомих публікацій [2], у яких зачіпаються проблеми ефективної роботи пошукових роботів, показує, що будь-який спеціалізований пошуковий робот повинен вирішувати наступні основні проблеми:

- побудова спеціалізованого алгоритму обходу веб-простору;

- відновлення інформаційних веб-ресурсів у загальній колекції документів, зібраної веб-краулером;
- облік різновидів форматів веб-ресурсів; організація інформаційного пошуку по схованому вебу;
- мінімізація навантаження на інформаційне джерело;
- розпаралелювання та масштабування процесу збору інформаційних веб-ресурсів; мінімізація навантаження на канали зв'язку у вебпросторі.

Для їхнього розв'язку розроблена та реалізована архітектура узагальненого ядра (рис. 1), яка може застосовуватися для створення будь-яких типів веб-краулерів. В основі цієї архітектури лежить краулер-процес. За структурою краулер-процес є класом, що містять поточну інформацію про стан пошукового робота (поточний список проаналізованих гіперпосилань, уміст веб-сторінок, кеш адрес веб-сторінок, список значимих джерел інформації), а також набір методів для здійснення пошуку та відновлення інформації в індексі. Дані методи дозволяють запускати та зупиняти процес пошуку.



Рисунок 1 – Зв'язок узагальненого ядра із зовнішніми модулями

Під процесом пошуку в рамках розглянутої архітектури розуміється ітеративний процес, що зупиняється при виконанні заданих умов (наприклад, число пройдених ітерацій, кількість отриманих гіперпосилань, збіжність результатів), які залежать від налаштувань конфігурації. Кожна ітерація даного процесу полягає в послідовному виконанні наступних основних модулів (рис. 1).

Багатопотоковий завантажувач веб-сторінок. Даний модуль виконує роль завантажника вебсторінок із сервера, на якому вони розташовані у веб-просторі. Також він є менеджером безпеки, який контролює кількість потоків, які виділяються для завантаження всіх інформаційних джерел у рамках однієї ітерації, і не оброблюють веб-ресурси, час відгуку яких

перевищує заданий. Усі параметри менеджера завантажень, налаштовуються, за замовчуванням число потоків рівно 10, а час відгуку - 2 с.

Модуль витягу гіперпосилань (Urls extraction) відшукує дочірні елементи із заданого набору веб-сторінок. Пошук таких елементів здійснюється шляхом витягу всіх гіперпосилань, що перебувають у початковій множині веб-сторінок, і їх додавання в чергу гіперпосилань. Процес обробки елементів черги виконується за допомогою набору паралельних і синхронізованих між собою потоків, які незалежно витягають посилання та додають результати обробки в загальну колекцію отриманих дочірніх веб-сторінок.

Модуль синтаксичного аналізу веб-сторінок (HTML parser) - витягає текст із завантажених веб-сторінок і визначає його кодування.

Нормалізація гіперпосилань (Urls normalization). Після того як з початкової множини веб-сторінок витягнуті всі гіперпосилання, необхідно відфільтрувати їхній адреси. За дану операцію в узагальненому ядрі відповідає модуль фільтрації адрес для веб-сторінок. Основне завдання такого модуля полягає в приведенні адреси кожного веб-ресурсу до стандартизованого виду за заданими критеріями (наприклад, ім'я хоста та протоколу кожного ресурсу повинно складатися із символів нижнього регістру або адреси всіх сторінок повинні приводитися до однакової канонічної форми, тобто закінчуватися символом «/»). Інакше кажучи, модуль фільтрації відсіває веб-сторінки з некоректними та дубльованими адресами, тим самим поліпшуючи продуктивність роботи пошукового робота.

Модуль кешування. Операція витягу вмісту веб-сторінок затрачає ресурси системи (інтенсивність витрат ресурсів системи залежить від росту числа веб-сторінок), що значним образом впливає на її продуктивність.

Тому була розроблена система кешування, призначена для прискорення процесу витягу гіперпосилань із веб-сторінок за рахунок повторного використання раніше витягнутого контенту. Зберігання контенту триває доти, поки не перевищений припустимий поріг пам'яті комп'ютера, використовуюваної додатком. Даний поріг додаток визначає динамічно, виходячи з наявного обсягу та вимог по використанню ресурсів комп'ютера, а також продуктивності. При перевищенні порога система звільняє пам'ять шляхом видалення з кешу контенту веб-сторінок, які рідше всього використовуються.

Колекція знайдених документів. Даний модуль зберігає інформацію про всі веб-ресурси і їх гіперпосилальних структурах, отриманих веб-краулером на кожній ітерації краулерпроцеса, і надає її користувачеві для подальших вебметрических досліджень.

Дана архітектура узагальненого ядра надає можливість настроювання інтеграції та інтерфейси, використання яких суттєво спрощує процес і мінімізує час додавання в архітектуру нових модулів, для створення потенційних зовнішніх модулів (нові алгоритм обходу, синтаксичний і

семантичний аналізатори та ін.).

Описана реалізація пошукового робота є базовим інструментом для вебметричних досліджень, тому на її основі можуть створюватися веб-краулери зі спеціалізованими алгоритмами обходу веба.

Архітектурні особливості узагальненого ядра. Перед експериментом у даних дослідженнях проводилася оцінка архітектурних можливостей узагальненого ядра та закордонних аналогів веб-краулерів (Heritrix, Openwebspider, Methanol Web Crawler) по наявності наступної функціональності:

- можливості багатопотокового завантаження вебсторінок;
- масштабування процесу збору веб-сторінок (на програмно-апаратному рівні);
- мінімізації навантаження на інформаційні веб-ресурси;
- мінімізації навантаження на канали зв'язку; гнучкості архітектури (можливості додавати нові модулі, алгоритми обходу); можливості спеціалізованого відновлення інформаційних веб-ресурсів в індексі

Таким чином, з поставленого експерименту випливає, що веб-краулер з узагальненим ядром ефективніше збирає та обробляє інформацію з веб-ресурсів, чим аналоги, які сильно уступають по продуктивності, гнучкості та масштабованості архітектури, а також по пристосованості до обробки інформації в вебі.

Надалі створений веб-краулер з узагальненим ядром планується використовувати в прикладних дослідженнях з підвищення вебметричних рейтингів університетів, наукових інститутів і інших наукових організацій

Література

1. Печников, А.А. Исследование связности научно-образовательного Веба / А.А. Печников, А.В. Чирков, Ю.В. Чуйко // Ученые записки Петрозаводского государственного университета. Естественные и технические науки. - 2011. -№ 8 (121). -С. 111-113.
2. Arasu, A. Searching the web [Text] / A. Arasu, J. Cho, H. Garcia-Molina [et al.] // ACM Transactions on Internet Technology. -Aug. 2001. -Vol. 1. -№ 1. -P. 2-43.
3. Bar-Ilan, J. How much information the search engines disclose on links to a web page? A case study of the «Cybermetrics» home page [Text] / J. Bar-Ilan // Proc.VIII Intern.Conf.on Scientometrics and Informetrics. - 2001. -Vol. 1. - P.63-73.
4. Raghavan, S. Crawling the Hidden Web / S. Raghavan, H. Garcia-Molina. -San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001. - P. 129-138.

Наукове видання

«Інтелектуальний потенціал – 2018» - збірник наукових праць молодих науковців і студентів з нагоди 30-річчя підготовки ІТ-фахівців в ХНУ/Колектив авторів – Хмельницький: ПВНЗ УЕП, 2018. – Ч.1: Комп'ютерні науки та інформаційні технології проектування. – 132 с.

**Відповідальність за зміст текстів і якість редагування матеріалів
покладена на авторів і наукових керівників.**

Комп'ютерна верстка: Чешун В.М.
Дизайн: Муляр І.В.

Здано до складання 5.11.18. Підписано до друку 10.11.18. Формат 60x84/16. Папір друкарський. Тираж 50 прим. Умовних друківаних аркушів – 8,25. Обліково-видавничих аркушів –1,85.

Редакційний відділ ПВНЗ УЕП 29016, м. Хмельницький, вул. Львівське шосе, 51/2.

ББК 74.480.278
С.88