

Кунгурцев О.Б.

Державний університет «Одеська політехніка»

Новикова Н.О.

Одеський національний морський університет

ПОБУДОВА МОДЕЛІ ПРЕДМЕТНОЇ ГАЛУЗІ В ПРОЦЕСІ ПРОЄКТУВАННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ

Розглянуто варіанти побудови моделі предметної галузі. Показано, що формування моделі зазвичай відбувається після виявлення вимог на пізніх етапах формування автоматизованих систем. Запропоновано будувати модель предметної галузі організації, для якої розробляється інформаційна система, на етапі формування вимог до інформаційної системи. Показано, що з цією метою можна використовувати документи, що підлягають аналізу для вироблення вимог. Встановлено, що основою моделі предметної галузі може бути словник предметної галузі. Проведено аналіз даних, що використовуються в процесі формування вимог до інформаційної системи, яка розробляється. Встановлено, що для побудови моделі предметної галузі можна використовувати додаткову інформацію про терміни (дії, посилання на документи) та документи (зв'язок зі структурою організації, функція).

Запропоновано розширений словник предметної галузі за рахунок введення понять «дія» та «об'єкт дії», встановлення зв'язку між термінами, що виконують функції дії та об'єкта дії, а також визначення посилань на документи, в яких ці терміни використовуються.

Створено математичну модель предметної галузі, що поєднує розширений словник предметної галузі та довідник документів. Для отримання довідника вибрані для аналізу документи проходять кластеризацію. У довіднику міститься інформація про тему документа, функції та підрозділи. Математична модель дозволяє об'єднувати документи в тематичні кластери, формувати відповіді на запити щодо термінів та документів. Для термінів можна визначити їхнє тлумачення, функції у системі та взаємодію, а також приналежність до документів. Для документів є можливість організувати пошук за темами, за термінами, що входять до них, за функціями та підрозділами. Також є можливість ранжування документів за рівнем відповідності до запиту. Виконано апробацію та обговорення запропонованих рішень. Результати апробації показали суттєве скорочення часу отримання інформації про предметну галузь у межах запропонованої моделі.

Ключові слова: словник предметної галузі, термін, модель, предметна галузь, об'єкт.

Постановка проблеми. Розробка інформаційних систем (ІС) починається з аналізу предметної галузі (ПГ) організації. Це надзвичайно трудомісткий та відповідальний етап виявлення вимог, що виконується висококваліфікованим спеціалістом – системним аналітиком. У процесі спілкування із замовником системний аналітик має зрозуміти його мову. Для цього складаються словники предметних галузей (СПГ) [1], з яких можна сформувати тезауруси. Джерелами інформації для побудови СПО в основному є документи, що використовуються в організації замовника. У процесі аналізу вимог до ІС також вивчається структура та розподіл обов'язків у організації замовника. Отримана інформація шляхом її структуризації та накопичення може бути представлена як модель ПГ та бути основою для побудови онтології ПГ організації замовника [2]. Питання про важливість та принципи використання онтологій, побу-

дованих на основі аналізу документів, розглядається в [3].

Модель онтології заведено представляти у вигляді [4]

$$O = \langle T, R, F \rangle,$$

де T – кінцева множина понять (концептів) предметної галузі,

– R – кінцева множина відносин між поняттями,

– F – кінцева множина функцій інтерпретації, заданих на концептах та/або відносінах.

У процесі розроблення вимог до ІС є можливість визначити T у вигляді термінів предметної галузі та деяких їх властивостей, а також частково визначити R і F у вигляді зв'язків термінів через документи та організаційну структуру.

Аналіз останніх досліджень і публікацій. Аналіз літературних даних мав на меті знайти рішення, пов'язані з побудовою моделі ПГ або її

онтології на початкових етапах проектування ІС, паралельно з виявленням вимог до ІС. Це дозволить суттєво скоротити витрати на побудову моделі ПГ.

У роботі [4] зазначено, що побудова онтології потребує великих ресурсів та часу, запропоновано автоматизувати цей процес. Однак вихідними даними для побудови мають бути ретельно підібрані та класифіковані документи, що важко виконати на етапі збирання та аналізу вимог. У роботі [5] розглянуто побудову універсального словника на основі аналізу документів для предметної галузі наукової літератури. Завдання спрощене завдяки використанню анотацій для класифікації документів. У статті [6] запропоновано приписувати документам додаткові характеристики, що визначаються середовищем їх використання та значенням, що буде використано у цій роботі. У роботі [7] запропоновано напівавтоматичний метод побудови онтології. Однак він спеціалізований для аналізу текстових даних Twitter та предметної галузі тероризму. Проблема зв'язку онтологій розглянута в [8]. Проте запропонована технологія передбачає лише окремий випадок – автоматизацію створення онтології предметної галузі шляхом перетворення заздалегідь отриманих файлів у форматі Excel у файли OWL. У роботі [9] пропонується вводити спеціальні формати для документів, що зберігаються у великих базах даних як з метою побудови онтологій, так і з метою виконання запитів. Таке рішення являє інтерес для ІС на стадії їх експлуатації, але не на стадії розробки. У роботах [10; 11] запропоновано на основі текстів будувати семантичні мережі як основу для онтологій ПГ. Очевидно, що таке рішення хоч і дозволить отримати якісний результат, але пов'язане із залученням значних ресурсів, що є неприйнятним на стадії побудови ІС.

ПГ організації, для якої створюється ІС, практично завжди неоднорідна, отже, аналізовані документи мали би бути класифіковані. Оскільки класифікація за змістом далеко не завжди виконується в організаціях [12], має сенс використовувати попередню кластеризацію [13], застосувавши її до всіх, а не лише до коротких документів.

Відомо, що складання словника предметної галузі є трудомістким процесом, що вимагає спеціальних знань. Тому має сенс скористатися наявними рішеннями [14], в основу якого покладено частотний аналіз корпусу документів.

На підставі аналізу літератури можна сформулювати таку проблему. Є необхідність інформацію, що одержана у процесі збору вимог до проєктованої ІС, доповнити, структурувати та

інтерпретувати таким чином, щоб отримати модель ПГ, яка має самостійне застосування та придатна для створення онтології ПГ.

Мета та завдання дослідження. Метою дослідження є скорочення часу та ресурсів на побудову моделі ПГ у процесі розробки ІС.

Для досягнення поставленої мети було сформульовано такі завдання:

- визначити необхідні з погляду побудови моделі ПГ дані та артефакти, які використовуються в процесі аналізу вимог до ІС, що проєктується;
- розширити визначення терміна з ПГ шляхом відображення дій;
- створити математичну модель ПГ;
- виконати апробацію запропонованих рішень.

Побудова моделі предметної галузі.

Ефективне використання даних, що одержуються в процесі аналізу вимог до проєктованої ІС. Одне з перших завдань, що вирішуються у процесі формулювання вимог, – побудова словника предметної галузі (СПГ). Словник є основою спілкування розробника і замовника, побудови інтерфейсів користувачів, складання інструкцій. Зазвичай словник складається на підставі аналізу документів, що характеризують ПГ замовника, і може бути поданий у такому вигляді [14]:

$$Dv = \langle Fw, mDe \rangle, \quad (1)$$

- де Dv – словник;
- Fw – передмова;
- mDe – множина словникових статей.

Кожна стаття словника може бути представлена кортежем:

$$De = \langle Tr, mTs, Tin \rangle, \quad (2)$$

де Tr – текст терміна та його ідентифікатор у словнику. У загальному випадку термін може бути багатослівним;

- mTs – множина синонімів терміна;
- Tin – дефініція (тлумачення) терміна.

Подання De досить для вирішення завдань аналізу вимог. Проте системний аналітик спільно з експертом ПГ у процесі побудови СПГ виконує значно більший обсяг робіт, ніж це потрібно для побудови словника. Крім цього, з документів, які обробляються для словника, можна отримати додаткову інформацію, корисну для моделі ПГ. Визначимо цю інформацію та джерела її отримання.

– Кожен термін може бути визначений як дія або об'єкт дії.

– Кожен термін може бути забезпечений множиною посилань на документи, у яких він виявлений.

– Документи однієї організації належать до різних тем (підпорядків), наприклад, до основної діяльності, управління кадрами, матеріально-технічного постачання тощо, отже, документи перед аналізом проходять класифікацію або кластеризацію [13].

– Документи, що створюються в процесі аналізу вимог, які також використовуються для побудови СПГ (опис варіантів використання, бізнес вимог тощо), містять функції, які має вирішувати ІС.

– Документи пов’язані з підрозділами організації замовника, де вони створюються та використовуються.

Математична модель предметної галузі.

Проведений аналіз раніше не використаних джерел інформації дозволяє подати математичну модель ПГ у вигляді двох компонентів: розширеного словника *DvExt*

$$DvExt = \langle Fw, mDeExt \rangle, \quad (3)$$

де *mDeExt* – множина розширених словникових статей,

та довідника документів *Dd*

$$Dd = \{mDf\},$$

де *mDf* = $\langle Did, Dc, mDf, mDep \rangle$ є довідкою про документ, яка містить

- *Did* – ідентифікатор документа;
- *Dc* – кластер, до якого належить документ;
- *mDf* – функція, яка пов’язана із документом;
- *mDep* – множина підрозділів, що використовують документ.

Зв’язок між словником *DvExt* та довідником *Dd* підтримується посиланнями статей словника на документи.

Представимо розширену статтю СПГ у вигляді:

$$DeExt = \langle Tr, mTs, Tin, Type, mA, mOb, mRDid \rangle \quad (4)$$

де *Tr* – термін;

– *Type* – тип терміна; якщо *Tr* – термін-дія, то *Type* приймає значення *a* (action), в іншому випадку – *o* (object);

– *mA* – множина дій, у яких бере участь термін *Tr*;

– *mOb* – множина об’єктів, пов’язаних з діями з множини *mA*;

– *mRDid* – множина посилань на документи, в яких було виявлено.

Пропонується така методика визначення статті розширеного словника *DvExt*.

1. Визначення типу об’єкта. Процедура виявлення типу об’єкта ґрунтується на використанні тлумачного словника української або іншої мови [23].

Представимо статтю тлумачного словника у вигляді:

$$DeG = \langle C, mTin, mA \rangle \quad (5)$$

де *C* – поняття, що визначається,

mTin – множина його дефініцій;

mAttr – множина атрибутів визначення поняття, таких як рід, закінчення, приклад використання та ін. Серед цих атрибутів для віддіслівних іменників повинна бути фраза виду «дія за значенням дієслова...». Позначимо цей атрибут *V* (verb).

Оскільки термін *Tr* у словнику *Dv* у загальному випадку є словосполучення, то на такому етапі зручно буде його уявити як множину слів – $Tr \Rightarrow mW$. Тоді умову виявлення дії можна представити у вигляді:

$$Cond = C \in mW \wedge V \in mA. \quad (6)$$

Якщо $Cond = true$, то *Type* приймає значення *a*, в іншому випадку – *o*.

2. Множина синонімів заповнюється термінами *mTs* з відповідного запису *De*.

3. Визначення множини дій і множини об’єктів, пов’язаних з діями з множини *mA*. Тут слід передбачити різні операції для термінів-дій та термінів-об’єктів.

Для термінів-дій до множини дій записується сам термін: $mA = \{Tr\}$.

Далі слід визначити множини об’єктів, пов’язаних з діями з множини *mA*.

Попередньо проаналізуємо терміни-дії та їхні синоніми, які містять кілька слів. Введемо операцію \in_A визначення належності терміна-об’єкта терміну-дії:

$$Tr_{OBJ} \in_A Tr_{ACT}, \quad (7)$$

де Tr_{OBJ} – термін-об’єкт, Tr_{ACT} – термін-дія. Результатом виконання операції буде *true*, якщо термін-об’єкт входить у термін-дія, і *false* у протилежному випадку. Наприклад, якщо термін-дія визначений словосполученням «розрахунок заробітної плати», а термін-об’єкт – «заробітна плата», то термін-об’єкт належить терміну-дії.

На підставі сказаного можна визначити множину mA' для деякого об’єкта-дії Tr_{ACT}

$$mA' = \{Tr_{OBJ} \mid Tr_{OBJ} \in_A Tr_{ACT} \vee Tr_{OBJ} \in_A mTs_j \wedge mTs_j \in mTs\} \quad j=1, ns. \quad (8)$$

Крім цього, оскільки термін-дія виступає у ролі підмета в реченні, а серед решти членів речення можуть зустрічатися терміни-об’єкти, то має сенс доповнити *mOb* усіма термінами-об’єктами з такого речення. Для цього потрібний другий прохід синтаксичного аналізатора, за якого для кожного речення *Offer* повинна виконуватися операція визначення елементів множини $mA'' = \{Tr_{OBJ} \mid \exists Tr_{ACT} = subject \wedge Tr_{OBJ} \in Offer\}$.

Наприклад, якщо як терміни раніше виділено «співробітник» та «прибутковий податок», то в результаті аналізу речення «Розрахунок заробіт-

ної плати співробітників» передбачає облік прибуткового податку та інших відрахувань, до множини mA будуть додані терміни «співробітник» та «прибутковий податок».

Після завершення другого проходу можна визначити остаточний варіант множини mA

$$mA = mA' \cup mA'' \quad (9)$$

Операції з виявлення дій для термінів-об'єктів можна виконати під час другого проходу синтаксичного аналізатора паралельно до обчислення множини mA для термінів-дій.

Для кожного речення *Offer* повинні виконуватись дві операції. Перша – визначає дію, пов'язану з терміном-об'єктом

$$mA = \{predicate_i \mid T_{Obj} = subject \wedge T_{Obj} \in Offer, \wedge (\exists T_{Obj} \in Offer \wedge T_{Obj}) \neq predicate\} \quad (10)$$

де $predicate_i$ є присудком у реченні, де термін-об'єкт грає роль підмета. Речення повинне містити інші терміни, окрім терміна у ролі підмета. Індекс при присудку і терміні вводится для забезпечення можливості поставити у відповідність присудку терміни, які беруть участь у такій дії.

Друга операція визначає терміни-об'єкти, які беруть участь у такій дії, але не є підметом у реченні.

$$mOb = \{T_{Obj} \mid T_{Obj} \neq subject \wedge T_{Obj} \in Offer\} \quad (11)$$

У цьому дослідженні ми обмежилися лише виявленням дій, у яких може брати участь термін-об'єкт. Виявлення інших атрибутів та умов виконання дій – предмет подальшого дослідження.

Крім дії рівня терміна, всі об'єкти через посилення на документ $mRDid$ можуть бути визначені як елементи дії рівня певного документа Did_i . Наприклад, усі терміни документа «*Опис варіанта використання. Замовлення обладнання*» можна віднести до дії «замовлення обладнання».

Можливі способи та сфери застосування моделі предметної галузі. У процесі розробки нового програмного продукту «під замовлення» складається тезаурус, що дозволяє замовнику та виконавцю проекту досягти однакового розуміння термінів із предметної галузі. Запропонована модель ПГ повністю виконує завдання тезаурусу і додатково, визначаючи дії над об'єктами, дозволяє визначити функції елементів ПГ.

Моделі ПГ можна використовувати на етапі уточнення функціональних вимог до системи, що розробляється, зокрема, у разі опису прецедентів [15]. Оскільки самі назви прецедентів рекомендується починати з віддієслівних іменників («реєстрація», «замовлення», «оформлення», ...), то в моделі ПГ можна визначити об'єкти, які використовуються у відповідних операціях.

У разі побудови моделі реляційної бази даних (РБД) часто використовують поняття «сутність» і «зв'язок» [16]. У моделі ПГ терміни є суттю, а дії фактично визначають зв'язки. Наявність або відсутність зв'язків по дії між термінами визначає наявність або відсутність зв'язків у РБД.

Одним зі способів інтеграції баз даних є побудова об'єктних моделей, що інтегруються РБД [17]. У цьому разі модель ПГ може бути побудована не на основі вихідних текстових документів, а на підставі конкретного змісту та зв'язків РБД. У такому вигляді модель ПГ може надати максимум інформації щодо оцінки ідентичності сутностей із різних РБД.

Натепер для багатьох ІС бажано створювати інтерфейс природною мовою, що істотно збільшує кількість користувачів системи. Модель ПГ може бути основою для побудови таких інтерфейсів, оскільки обмежує набір термінів, зрозумілих для ІС з точки зору змісту інформації, що зберігається, і дій над нею.

Запропонована модель може бути основою для виконання різних запитів щодо пошуку документів у корпоративній системі [18] шляхом виявлення у запиті термінів і дій та зіставлення їх з інформацією, що зберігається *DvExt*.

Апробація та обговорення результатів дослідження. Для побудови моделі предметної галузі на підставі аналізу текстових документів було проаналізовано предметну галузь «Електрична сфера в Україні». Для експерименту було взято такі документи: «Про затвердження Правил користування електричною енергією для населення» [19], «Інструкція зі складання актів про порушення правил користування електричною енергією для населення» [20], документи містили 9760 слів.

Експеримент проводився в операційному середовищі Windows 10, мова програмування – java. Для виділення термінів використовувалося програмне забезпечення для створення тлумачного словника предметної галузі [14]. Для визначення дефініцій виділених термінів використовувався програмний продукт Dictionary Of Interpretations [21]. Для виконання синтаксичного аналізу використали синтаксичний аналізатор [22]. Результати експерименту були представлені у XML файлі.

На підставі аналізу відібраних документів було отримано словник для предметної галузі «Електрична сфера в Україні». Фрагмент словника представлений у табл. 1.

На підставі аналізу термінів, що увійшли до словника предметної галузі, було створено роз-

ширений словник предметної галузі, фрагмент якого представлений у таблиці 2.

На підставі аналізу записів у табл. 2 в автоматизованому режимі зроблено такі твердження, що характеризують предметну галузь (слова в дужках внесені експертом).

«Енергопостачальник укладає угоду».

«Енергопостачальник поставляє електроенергію».

«Енергопостачальник припиняє (подачу) електроенергії».

«Споживач повинен укласти угоду».

«Споживач може внести (пропозиції до) угоди».

«Споживач сплачує за електроенергію».

«Оплата послуг здійснюється (за) підключення».

«Оплата послуг здійснюється (за) ремонт».

Відповідно до методики, запропонованої в роботі [23], було визначено час складання розширеного словника ПГ для поданих документів в автоматизованому режимі та в режимі ручної роботи. У першому випадку воно становило 1 годину (з урахуванням часу редагування твер-

джен аналітиком), у другому – 12,5 години. Таким чином, було зафіксовано скорочення часу аналізу ПГ більш ніж у 10 разів.

У проведеному експерименті було використано документи з Інтернету. У подальших експериментах планується розширити предметну галузь додатковою інформацією про документи, наприклад, датою створення, типом тощо, а також додатковою інформацією про структуру організації-замовника ІС. Також планується сформувати мову запитів до моделі ПГ.

Висновки. Проведено аналіз даних, що використовуються в процесі формування вимог до ІС, яка розробляється. Встановлено, що для побудови моделі ПГ можна використовувати додаткову інформацію про терміни (дії, посилання на документи) та документи (зв'язок зі структурою організації, функція).

Запропоновано розширений СПГ за рахунок введення понять «дія» та «об'єкт дії», встановлення зв'язку між термінами, визначення посилань на документи.

Таблиця 1

Фрагмент словника предметної галузі

Термін <i>Tr</i>	Частота входжень <i>F</i>	Синоніми <i>mTs</i>	Визначення <i>Tin</i> з тлумачного словника
Енергопостачальник	0,0217	продавець	учасник оптового ринку електричної енергії, який купує електричну енергію з метою продажу її споживачам
Електроенергія	0,0341	енергія	енергія, що передається електричним струмом
Споживач електричної енергії	0,0203	замовник, покупець	громадянин або організація, яка замовляє, отримує товари, послуги, роботи

Таблиця 2

Фрагмент розширеного словника предметної галузі

Термін <i>N</i>	Тип <i>Type</i>	Синоніми <i>Ns</i>	Дія <i>A</i>	Термін <i>Na</i>	Тлумачення <i>T</i>
Енергопостачальник	obj	продавець	припиняє, укладає, постачає	договір, електроенергія	учасник оптового ринку електричної енергії, що купує електричну енергію з метою продажу її споживачам
Електроенергія	obj	енергія	визначається, оплачується	споживач	енергія, що передається електричним струмом
Споживач електричної енергії	obj	замовник, покупець	може внести (пропозиції), повинен укласти, сплачує	оплата, угода	громадянин або організація, яка замовляє, отримує товари, послуги, роботи
Оплата послуг	act	розрахунок	здійснюється, зараховується	Підключення, ремонт	розрахунок з організацією-постачальником за витрачену електроенергію, за виконання проектних, ремонтних та інших робіт
Вимкнення електроенергії	act	роз'єднання, вимкнення	здійснюється	споживач	тимчасове відключення електропостачання великих груп споживачів постачальником

Створено математичну модель ПГ, що поєднує розширений СПГ та довідник документів, що містить інформацію про тему документа, функції та підрозділи. Математична модель дозволяє давати відповіді на запити, на які

раніше неможливо було відповісти, або відповідь вимагала багато часу.

Виконано апробацію запропонованих рішень. Результати показали суттєве скорочення часу отримання інформації про ПГ.

Список літератури:

1. Wiegers K., Beatty J. Software Requirements. 3rd ed. Redmond, WA : Microsoft Press, 2013. 637 p. ISBN 978-0-7356-7966-5.
2. Цуканова Н.И. Онтологическая модель представления и организации знаний : учебное пособие для вузов. 2015 г. 272 с. ISBN 978-5-9912-0454-5 http://www.techbook.ru/book.php?id_book=759.
3. Быстров И.И., Тарасов Б.В., Хорошилов А.А., Радоманов С.И. Основы применения онтологии и компьютерной лингвистики при проектировании перспективных автоматизированных информационных систем. *Системы и средства информатики*, 2015. Том 25, выпуск 4. С. 128–149. DOI: <https://doi.org/10.14357/08696527150410>.
4. Куркин А.Н. Исследование методов автоматической генерации онтологии на основе структурированного материала. *Современная техника и технологии*, 2016. № 10. URL: <https://technology.snauka.ru/2016/10/10842>.
5. Constantin A., Peroni S., Pettifer S., Shotton D., Vitali F. The Document Components Ontology (DoCO). *Semantic Web*. 2016. Vol. 7, No 2. P. 167–181. DOI: 10.3233/SW-150177.
6. Furner J. The Ontology of Documents, Revisited. *Proceedings from the Document Academy*. 2019. Vol. 6. Iss. 1 , Article 1. DOI: 10.35492/docam/6/1/114 <https://ideaexchange.uakron.edu/docam/vol6/iss1/1>.
7. Jinda R., Seeja K.R., Jain S. Construction of domain ontology utilizing formal concept analysis and social media analytics. *International Journal of Cognitive Computing in Engineering*. 2020. Vol. 1. P. 62–69. DOI: 10.1016/j.ijcce.2020.11.003.
8. ElAsri H., Jebbor F., Benhlima L. Building a Domain Ontology for the Construction Industry. *Towards Knowledge Sharing International Conference on Digital Technologies and Applications ICDTA*. 2021. P. 1061–1071.
9. Lembo D., Scafoglieri F. M. Ontology-based Document Spanning Systems for Information Extraction. *International Journal of Semantic Computing*. 2020. Vol. 14, No. 01. P. 3–26. DOI:10.1142/S1793351X20400012.
10. McCrae J.P., Bosque-Gil J., Gracia J., Buitelaar P., Cimiano Ph. OntoLex-Lemon Model: development and applications. URL: <http://john.mccr.ae/papers/mccrae2017ontolex.pdf> (дата звернення: 21.11.2020).
11. Горшков С., Кондратьев К., Шебалов Р. Онтологии: от текста к фактам. URL: <https://www.osp.ru/os/2020/04/13055699> (дата звернення: 06.11.2020).
12. Плешкевич А.А., Королев И.Д. Классификация документов в системе электронного документооборота. *Молодой ученый*. 2018. № 10 (196). С. 21–24. URL: <https://moluch.ru/archive/196/48705/>.
13. Kungurtsev O., Zinovatna S., Potochniak Ia., Novikova N. Development of methods for pre-clustering and virtual merging of short documents for building domain dictionaries. *Eastern-European Journal of Enterprise Technologies*. 2020. Vol. 5, No 2(107). P. 39–47. DOI: <https://doi.org/10.15587/1729-4061.2020.215190>.
14. Kungurtsev O., Zinovatnaya S., Potochniak Ia., Kutasevych M. Development of information technology of term extraction from documents in natural language. *Eastern-European Journal of Enterprise Technologies*. 2018. Vol. 6, No. 2 (96). P. 44–51. DOI: <https://doi.org/10.15587/1729-4061.2018.147978>.
15. Kungurtsev O., Novikova N., Reshetnyak M., Cherepinina Ya., Gromaszek K., Jarykbassov D. Method for defining conceptual classes in the description of use cases. *Proc. SPIE 11176, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments*. 2019. 1117624. DOI: 10.1117/12.2537070.
16. Hernandez M.J. Database Design for Mere Mortals: A Hands-On Guide to Relational Database Design. 3rd edition. Addison-Wesley, 2013. 680 p.
17. Хоп Г., Вульф Б. Шаблоны интеграции корпоративных приложений. Издательство «Диалектика», 2019. 672 с.
18. Ward S. How to Create a Document Management System. URL: <https://www.thebalancesmb.com/creating-a-document-management-system-2948084> (дата звернення: 22.05.2021).
19. Про затвердження Правил користування електричною енергією. URL: <http://zakon2.rada.gov.ua/laws/show/z0417-96>.
20. Инструкция по составлению актов о нарушении правил пользования электрической энергией для населения. URL: <http://leg.co.ua/instrukcii/energonadzor/instrukciya-po-sostavleniyu-aktov-o-narushenii-pravil-polzovaniya-elektricheskoy-energiyey-dlya-naseleniya.html>.
21. Kungurtsev O., Novikova N., Kozhushan M. Automation of Serching for Terms in the Explanatory Dictionary. *Proceedings of Odessa Polytechnic University*. Issue 3(62). 2020. P. 9–100.

22. Автоматическая Обработка текста. URL: <http://www.aot.ru/> (дата звернення: 2.03.2021).

23. Кунгурцев А.Б., Поточняк Я.В., Силяев Д.А. Метод автоматизированного построения толкового словаря предметной области. *Технологический аудит и резервы производства*. 2015. № 2(22). С. 58–63.

**Kungurtsev O.B., Novikova N.O. CONSTRUCTING OF A SUBJECT AREA MODEL
IN THE PROCESS OF THE INFORMATION SYSTEM DESIGNING**

The options for constructing of a subject area model are considered. It is shown that the process of a model forming usually occurs after the identification of requirements at the late stages of the automated systems formation. It is proposed to construct the subject area model of the organization, for which an information system is developed at the stage of forming the requirements for IS. It is shown that for this purpose it is possible to use documents which are to be analyzed to develop requirements. It is found that the vocabulary of the subject area can serve as the basis for the subject area model. The analysis of the data used in the process of forming the requirements for the developed IS is carried out. It is found that to construct a subject area model, it is possible to use additional information about terms (operations, links to documents) and documents (connection with the organization structure, function).

An extended vocabulary of the subject area is proposed by introducing the concepts of “operation” and “object of operation” and establishing connections between terms which perform functions of an operation and an object of operation, as well as defining references to documents in which these terms are used.

The mathematical model of the subject area, which combines an extended vocabulary of the subject area and a reference book of documents, is created. To obtain a reference book, the documents selected for analysis undergo clustering. The reference book contains information on the topic of the document, function and subunit. The mathematical model makes it possible to combine documents into thematic clusters, form responses to queries regarding terms and documents. There is a possibility for terms to define their interpretations, functions in the system and their interactions, as well as their belonging to documents. It is possible for documents to organize a search by topics, by terms included in them, by functions and subunits. It is also possible to rank documents in accordance with relevance to the request. Approbation and discussion of the proposed solutions are completed. The results of approbation showed a significant reduction in the time for obtaining information about the subject area within the scope of the proposed model.

Key words: *vocabulary of the subject area, term, model, subject area, object.*