

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
БУДІВНИЦТВА І АРХІТЕКТУРИ**

**ШАБАЛА ЄВГЕНІЯ ЄВГЕНІВНА**



УДК 004.896; 004.855.5

**ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ  
ІДЕНТИФІКАЦІЇ ЗМІСТУ БУДІВЕЛЬНИХ НОРМАТИВІВ**

05.13.06 – інформаційні технології

**АВТОРЕФЕРАТ**  
дисертації на здобуття наукового ступеня  
кандидата технічних наук

Київ 2015

Дисертацією є рукопис

Робота виконана на кафедрі інформаційних технологій Київського національного університету будівництва і архітектури Міністерства освіти і науки України

**Науковий керівник** – доктор технічних наук, професор  
**Тесля Юрій Миколайович**,  
Київський національний університет імені Тараса Шевченка Міністерства освіти і науки України,  
декан факультету інформаційних технологій, лауреат  
Державної премії України в галузі науки і техніки  
(м. Київ)

**Офіційні опоненти:** доктор технічних наук, професор,  
**Павленко Петро Миколайович**,  
Національний авіаційний університет Міністерства  
освіти і науки України,  
заступник директора з навчально-методичної роботи  
інституту інформаційно-діагностичних систем  
(м. Київ)

кандидат технічних наук, доцент  
**Данченко Олена Борисівна**,  
Університет економіки та права «КРОК» Міністерства  
освіти і науки України (м. Київ),  
в.о. завідувача кафедри бізнес адміністрування та  
управління проектами

Захист відбудеться «26» листопада 2015 р. о 12<sup>00</sup> годині на засіданні спеціалізованої вченої ради Д 26.056.01 при Київському національному університеті будівництва і архітектури за адресою: 03680, м. Київ, Повітрофлотський просп., 31., ауд. 466.

З дисертацією можна ознайомитись в бібліотеці при Київському національному університеті будівництва і архітектури за адресою: 03680, м. Київ, Повітрофлотський просп., 31

Автореферат розіслано «23» жовтня 2015 р.

Вчений секретар спеціалізованої  
вченої ради доктор технічних наук,  
професор



С.В. Цюцюра

## ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

**Актуальність теми.** В даний час різними сховищами знань накопичені величезні інформаційні масиви в будівельній галузі. Однак відсутність можливості оперативно отримати найбільш актуальну та повну інформацію по конкретній темі знецінює значну частину накопичених інформаційних ресурсів. Тому все більше наукових досліджень концентрується на розробці формальних моделей і методів обробки природної мови. Одним з найважливіших напрямків цієї проблематики є класифікація різних лінгвістичних одиниць (текстів, слів, словосполучень), що реалізується, практично у всіх застосуваннях лінгвістичного процесора. Саме успішність застосування сучасних методологій обробки природномовних текстів, є фактором, який визначає ефективність створення, обліку, управління і актуалізації документації будівельної галузі. Але більшість існуючих моделей або не дозволяють автоматизувати процедури класифікації, або дають високий рівень похибок і низьку точність при їх практичній реалізації. Це пов'язано з тим, що до теперішнього часу розробка систем автоматичної обробки текстів природною мовою відбувалася без використання смислового аналізу або з його мінімальним використанням. Таким чином виникла потреба у створенні інформаційної технології ідентифікації змісту текстів будівельної тематики, яка дозволить обробляти природномовні тексти будівельних нормативів.

Проведене в дисертаційній роботі дослідження ґрунтується на фундаментальних працях видатних вчених: М.М. Амосова, В.Ю. Крилова, Ю.І. Морозова, Р.В. Палагіна, Г.С. Поспелова, Ю.М. Теслі. В загально-філософському плані в контексті проблеми штучного інтелекту становлять інтерес праці І.С. Добронравова, П.В. Копніна, О.В. Палагіна, М.В. Поповича, В.С. Швирьова, А.П. Шептуліна, А.С. Щербакова.

Проблемами застосування штучного інтелекту в галузі будівництва, прогнозування, лінгвістичного аналізу займалися дослідники, такі як: Р. Браун, Г. Дженкінс, Ч. Хольт, Д. Бриллінджер, К. Верчелліс, Ю. П. Лукашин, А.Г. Івахненко, В.Д. Романенко, П.І. Бідюк, В.Є. Снитюк, Е.Є. Тихонов.

В основу дисертаційного дослідження покладено задачу створення інформаційної технології ідентифікації змісту текстів нормативних документів, яка може бути впроваджена зокрема у будівельній сфері. При розробці системи було запропоновано моделі та методи, які забезпечать скорочення часу на пошук необхідних підрозділів та оптимізують процес реферування текстів будівельних нормативів. До таких напрямків належать використання апарату теорії несилової взаємодії, штучного інтелекту, проектування кошторисів та іншої документації з будівельної галузі, лінгвістичний аналіз даних та реферування природномовних текстів.

Дослідження, які пов'язані з розробленням інформаційної технології ідентифікації змісту текстів нормативних документів в умовах постійних змін, доповнення існуючих та утвердження нових будівельних нормативів є актуальними і мають не тільки практичне значення для кошторисників-проектувальників, інвесторів та аналітиків, але й теоретичне: розробка концепцій і математичного

інструментарію, що може скласти основу для подальших наукових досліджень в цій галузі.

**Зв'язок роботи з науковими програмами, планами, темами.** Дисертаційна робота виконана в Київському національному університеті будівництва і архітектури в рамках наукових тем: «Створення системи перевірки ступеню унікальності наукових робіт» (номер державної реєстрації 0114U000126).

В межах цієї теми досліджені процеси автоматичного вилучення змісту текстів будівельних нормативів на основі аналізу навчальної вибірки сформованої експертом з будівельної галузі, розроблено ієрархію структури нормативів та виконано реферування текстів будівельних нормативів.

Результати наукових досліджень були впроваджені в науково-дослідний інститут будівельного виробництва (НДІВВ) Міністерства регіонального розвитку та будівництва України при формуванні нормативної документації на проектно-вишукувальні роботи та конструкторські розробки та у діяльності ТОВ «Аверсбуд».

**Об'єктом дослідження** є процеси ідентифікації змісту будівельних нормативів.

**Предметом дослідження** є інформаційна технологія та моделі ідентифікації змісту будівельних нормативів.

**Метою дисертаційного дослідження** є побудова інформаційної технології розпізнавання змісту та пошуку будівельних нормативів з наперед заданими властивостями, що дає можливість підсилення аналітичного апарату для проектних організацій.

**Задачами даного дослідження** є:

- аналіз методів і моделей управління інформацією в нормативній базі будівництва;
- розробка науково-методологічного та понятійного базису управління процесом лінгвістичного аналізу текстів будівельних нормативів;
- дослідження реакції зовнішніх впливів та впливу людського мислення на хід ідентифікації змісту природномовних текстів будівельних нормативів;
- розробка моделей формалізації змісту інформаційного масиву будівельних нормативів;
- використання рефлексивного підходу до побудови інформаційної технології ідентифікації змісту текстів будівельних нормативів;
- моделювання впливу людських факторів на процес детального визначення змісту будівельних нормативів;
- проведення експериментальних досліджень;
- розробка практичних інструментів ідентифікації змісту текстів будівельних нормативів.

**Методи дослідження.**

В процесі досліджень були використані методи, які ґрунтуються на використанні теорії систем, методів системного аналізу, теорії моделювання, об'єктно-орієнтованого програмування і теорії проектування та реалізації автоматизованих інформаційних систем, теорії несилового взаємодії з метою виявлення та класифікації інформаційних взаємодій, орієнтованих на формування

нових знань в системі на основі аналізу факторів впливу та побудови адекватної цим взаємодіям математичної моделі процесів ідентифікації змісту природномовних текстів на основі використання інтроформаційних методів. Були використані методи теорії ймовірностей, експертні методи для розробки моделей, інтелектуальної обробки даних і теорії прийняття рішень для розробки інформаційної технології ідентифікації змісту текстів будівельних нормативів.

**Наукова новизна дисертаційного дослідження** обумовлена тим, що вперше вирішена науково-технічна задача розробки рефлексорних моделей і методів лінгвістичного аналізу змісту текстів будівельних нормативів.

У дисертаційній роботі автором отримані такі результати:

- *вперше* для забезпечення розрахунку несилового впливу фрагментів речень на вибір ключових понять у нормативній базі в будівельній галузі запропоновано застосування інтроформаційного методу, що дозволяє формально підійти до оцінки впливу на визначення змісту нормативних документів;

- *вперше* описано процес визначення впевненості в правильності інтроформаційного змісту речення з нормативної бази, який пропонується реалізувати через оцінку сумісних умовних ймовірностей появи фрагментів тексту по частковим умовним ймовірностям, що дозволить створювати інформаційні технології ідентифікації змісту будівельних нормативів не на основі евристики, а використовуючи математичний апарат теорій імовірності та несилової взаємодії;

- *вперше* запропоновано визначення ключових понять тексту на основі рефлексорно-ймовірнісного підходу, що характеризується описом поведінки системи на основі рефлексів та дозволяє якісно обробляти нормативну документацію та скоротити час на пошук необхідних підрозділів будівельних нормативів;

- *удосконалено* міру впливу та методи оперування числовими значеннями впливів словосполучень речень на ідентифікацію змісту природномовних текстів через введення імовірнісних оцінок впливу, що відрізняється від існуючих введенням нечіткої міри інформаційної взаємодії та дозволяє оптимізувати дії по вибору ключових понять речень;

- *отримала подальший розвиток* математична модель оперування мірою впливу природномовних текстів на зміст будівельних нормативів, що відрізняється від існуючих орієнтацією не тільки на виділення впливових слів, але і на виділення впливових фрагментів речень, що дозволяє застосувати інтроформаційний метод до вирішення задачі ідентифікації змісту будівельних нормативів;

**Практичне значення одержаних результатів.** Розроблена інформаційна технологія ідентифікації змісту текстів нормативних документів, яка була створена на основі розроблених моделей та методів дала можливість використовувати її на практиці, зокрема в діяльності ТОВ «Аверсбуд» при розробці окремих практичних заходів щодо оперування інтроформацією в природних і штучних інтелектуальних системах. В ТОВ «Алісбуд» та ТОВ «Альфа-сервіс» ця технологія була застосована для підвищення ефективності процесу обробки нормативної документації.

**Особистий внесок здобувача.** Наукові положення, розробки та висновки дисертаційної роботи є результатом самостійно проведеного автором дослідження в області інтелектуального аналізу текстів будівельних нормативів. Внесок автора в праці, опубліковані у співавторстві, конкретизовано в списку публікацій.

**Апробація результатів дисертації.** Основні результати і дисертаційна робота в цілому апробована на 10 міжнародних наукових конференціях: Всеукраїнська науково-практична інтернет-конференція «Автоматизація та комп'ютерно-інтегровані технології у виробництві та освіті: стан, досягнення, перспективи розвитку» (м. Черкаси, 2013 р.); Всеукраїнська науково-практична інтернет-конференція «Автоматизація та комп'ютерно-інтегровані технології у виробництві та освіті: стан, досягнення, перспективи розвитку (м. Черкаси, 2014 р.)»; Наукова конференція молодих вчених, аспірантів і студентів Київського національного університету будівництва і архітектури: 2011р., 2012 р., 2013 р. Науково-практичні конференції Київського національного університету будівництва і архітектури: 72-я (2011 г.), 73-я (2012 р.), 74-я (2013 р.), Перша міжнародна науково-практична конференція «Управління розвитком технологій» (2014 р.), Друга міжнародна науково-практична конференція «Управління розвитком технологій» (2015 р.).

**Публікації.** Основні результати дисертаційної роботи опубліковані у 12-ти наукових працях, серед яких: у фахових виданнях – 5; зарубіжних статей – 1; в матеріалах конференцій – 6.

**Структура й обсяг роботи.** Дисертація складається із вступу, чотирьох розділів, висновків, списку використаних джерел зі 130 найменувань (розміщених на 19 сторінках), 3-х додатків (на 8 сторінках) та містить 7 таблиць і 22 рисунки. Загальний обсяг дисертаційної роботи - 170 сторінок. Основний текст викладений на 140 сторінках.

## ОСНОВНИЙ ЗМІСТ РОБОТИ

**У вступі** обґрунтовано актуальність теми дисертації, сформульовані її мета та завдання, об'єкт та предмет дослідження. Вказуються основні методи дослідження, наукова новизна та практичне значення отриманих результатів. Зазначений особистий внесок здобувача та наведено дані про реалізацію, апробацію та публікацію результатів досліджень.

**У першому розділі** наведено основні теоретичні відомості, які стосуються особливостей побудови і архітектури інформаційних технологій ідентифікації змісту будівельних нормативів, лінгвістичного аналізу, розробки та ведення нормативної бази, аналіз структури існуючої нормативної бази в будівництві. Також виявлені основні принципи розвитку нормативного забезпечення будівництва та проведений аналіз існуючих підходів до розробки інтелектуальних систем. Виявлено, що побудова інформаційної технології ідентифікації змісту є затребуваною в сфері будівництва, є одною зі складових успішного управління нормативною базою та підготовки процесів будівництва.

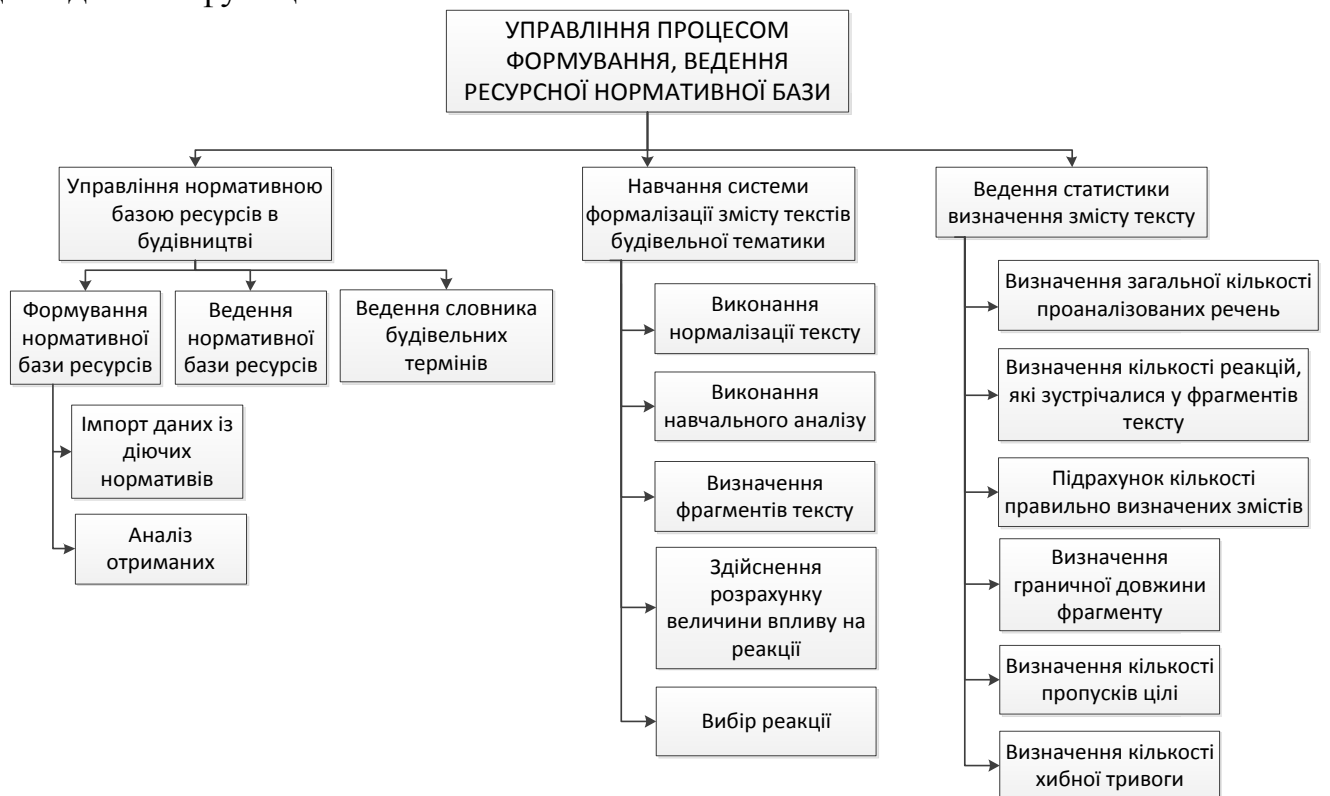
Визначено, що побудова та впровадження інформаційної технології

ідентифікації змісту будівельних нормативів є актуальним напрямком що може підвищити точність та якість визначення змісту нормативних документів, і як наслідок, скорити час на реферування природномовних текстів.

У другому розділі розглядається розробка науково-методологічного та понятійного базису управління процесом лінгвістичного аналізу текстів будівельних нормативів.

Показано, що основну увагу при проектуванні інформаційної технології ідентифікації змісту будівельних нормативів слід приділяти її здатності інтелектуальної обробки даних, а саме, можливості навчання інформаційної технології на основі досвіду її використання і можливості роботи із знаннями на основі формалізації тексту.

Для розробки інформаційної технології ідентифікації змісту будівельних нормативів було побудовано дерево функцій, яке представлено на рис. 1. Воно являє собою декомпозицію функцій системи і формується з метою детального дослідження функціональних можливостей системи.



**Рис. 1** Дерево функцій інформаційної технології ідентифікації змісту текстів будівельних нормативів

Для даного дослідження було використано рефлексорно-ймовірнісний підхід. Цей підхід характеризується описом поведінки системи на основі рефлексів. Рефлекс є основною формою діяльності нервової системи організму, це означає, що психічна діяльність регулюється рефлексорно. Рефлекси – реакції на внутрішні та зовнішні впливи. В основі рефлексів полягає наступний тезис: якщо це вже було, і деяка реакція була позитивно підкріплена, то необхідно зробити теж саме. При навчанні інтелектуальної системи спостерігаються такі ж рефлекси. Під реакцією слід розуміти правильне визначення змісту тексту. В процесі визначення змісту

нормативних документів вироблена реакція вже може і не бути правильною. Точніше, той же результат (оцінка того чи іншого змісту) може відповідати дійсності, а може і не відповідати.

Тоді можливості отримати очікуваний результат необхідно представити через певну ймовірність.

$$p_i \approx r_i / k_i, \quad (1)$$

де  $p_i$  – ймовірність того, що вироблений рефлекс дасть потрібний результат;  
 $r_i$  – кількість речень, в яких було правильно визначено зміст;  
 $k_i$  – загальна кількість речень.

Визначення змісту на основі оцінки результатів минулих дій можна представити в вигляді, який наведено в табл. 1.

Таблиця 1

Зміст	Оцінка результату визначення змісту
$E_1$	$G_1$
$E_2$	$G_2$
...	...
$E_i$	$G_i$
...	...
$E_m$	$G_m$

Тоді, при прийнятті рішень, повинен обиратися зміст із найвищою оцінкою правильності:

$$E_i = \{E_{i0} | E_{i0} \in E \wedge G_{i0} = \max_i(G_i)\}, \quad (2)$$

де  $E$  – множина варіантів змісту текстів будівельних нормативів;  
 $G_{i0}$  – оцінка результату визначення змісту  $|E_{i0} \in E$ .

Тоді можливість отримати очікуваний результат може бути охарактеризована ймовірністю (табл. 2).

Таблиця 2

Зміст	Скільки разів визначався зміст	Скільки разів був отриманий необхідний результат
$E_1$	$k_1$	$r_1$
$E_2$	$k_2$	$r_2$
...	...	...
$E_i$	$k_i$	$r_i$
...	...	...
$E_m$	$k_m$	$r_m$

Для оцінки ймовірності правильно визначеного змісту тексту нормативної бази в будівництві необхідно:

1. Текст будівельних нормативів розкласти на множину фрагментів:

$$C = \{c_{hg}\}, hg = \overline{1, f}, \quad (3)$$



де  $C$  – текст;  
 $c_{hg}$  – фрагмент тексту, який часто повторюється;  
 $f$  – кількість фрагментів.

2. По частоті успішних визначень змісту будівельних нормативів можна визначити приблизні значення ймовірності кожного із класів змісту за умови, що фрагмент нового тексту в цей момент був присутнім:

$$\forall E_j \in E, c_{hg} \in C : p(E_j / c_{hg}) \approx n(E_j / c_{hg}), \quad (4)$$

де  $n(E_j / c_{hg})$  – частота визначення змісту  $E_h$  за умови, що фрагмент  $c_{hg}$  в тексті був присутнім;

$p(E_j / c_{hg})$  – ймовірність визначення змісту  $E_h$  за умови, що фрагмент  $c_{hg}$  в тексті був присутнім.

3. По змісту  $E_j$  за його безумовною ймовірністю  $p(E_j)$  і частковими умовними ймовірностями  $p(E_j / c_{hg_1}), \dots, p(E_j / c_{hg_i}), \dots, p(E_j / c_{hg_f})$  необхідно оцінити сумісну умовну ймовірність:

$$p(E_j / c_{hg_1}, \dots, c_{hg_i}, \dots, c_{hg_f}) = p(E_j / C) \quad (5)$$

Зміст тексту представляється в якості продукту, який продукується інтелектуальним апаратом людини. Семантична складова тексту, яка є суттю і продуктом процесів несилової (інформаційної) взаємодії в інтелектуальному апараті людини, визначає його зміст.

Виділення фрагментів в тексті може відбуватися різними способами. Елементарним фрагментом можна вважати окремий символ, буквосполучення, склад, слово. Кількість символів у фрагментів тексту будемо називати довжиною фрагменту тексту. Графік залежності якості визначення змісту будівельних нормативів від довжини фрагментів тексту представлено на рис. 2.



**Рис. 2** Залежність ймовірності правильного визначення змісту будівельних нормативів від довжини фрагментів тексту

В роботі була запропонована числова міра впливу на характеристики ідентифікації змісту будівельних нормативів. В основі розрахунку міри впливу

правильно визначений зміст будівельних нормативів. Для формалізації цього значення запропоновано відслідковувати різноманітні впливи на процес ідентифікації змісту текстів, що дозволить отримати залежності відхилення автоматично визначеного змісту будівельних нормативів від змісту, визначеного експертом.

Методами теорії ймовірності вирішити задачу знаходження сумісної умовної ймовірності правильно визначеного змісту тексту будівельного нормативу, по частковим умовним і безумовній, не можна. Тому необхідно сформулювати задачу, як оцінити сумісну умовну ймовірність по частковим. Тобто застосувати такий метод оцінки сумісної умовної ймовірності по частковим, яка дозволить вибрати той зміст, який був би визначений і експертом. Тобто:

$$\forall E_h \in E \exists E_h \in E : p(\eta_M(E_h/C) \geq \eta_M(E_h/C) / p(E_h/C) \geq p(E_h/C)) \approx 1, \quad (6)$$

$$\forall E_j \in E, c_{hg} \in C : p(E_j/c_{hg}) \approx n(E_j/c_{hg}), \quad (7)$$

де  $\eta_M(E_h/C)$  – оцінка сумісної умовної ймовірності змісту  $e_h$  в тексті  $C$ , яка була отримана методом оцінки сумісної умовної ймовірності по частковим;

$\eta_M(E_h/C)$  – оцінка сумісної умовної ймовірності змісту  $e_h$  в тексті  $C$ , яка була отримана методом оцінки сумісної умовної ймовірності по частковим;

$p(\eta_M(E_h/C) \geq \eta_M(E_h/C) / p(E_h/C) \geq p(E_h/C))$  – умовна ймовірність того, що якщо умовна ймовірність визначення змісту  $E_h$  максимальна, то і оцінка сумісної умовної ймовірності змісту  $E_h$  максимальна.

Вираз (7) означає наступне, якщо сумісна умовна ймовірність визначення змісту  $e_h$  максимальна, то майже завжди максимальна її оцінка. Оптимальний метод  $M$  (де  $M$  – метод оцінки сумісної умовної ймовірності за частковими) завжди дає найвищу оцінку найбільшій сумісній умовній ймовірності. Для оцінки ефективності цього методу, значення  $p(E_h/C), p(E_h/c_{hg_1}), \dots, p(E_h/c_{hg_i}), \dots, p(E_h/c_{hg_f})$ , можна отримати дослідним шляхом. Відхилення виразу (7) від одиниці буде критерієм ефективності методу оцінки сумісних умовних ймовірностей по частковим:

$$1 - p(\eta_M(E_h/C) \geq \eta_M(E_h/C) / p(E_h/C) \geq p(E_h/C)) \rightarrow \min, \quad (8)$$

при обмеженнях:

$$1. C = \{c_{hg}\}, hg = \overline{1, f}.$$

$$2. E = \{E_h\}, h = \overline{1, m}.$$

$$3. \forall E_h \in E : p(E_h/c_{hg_1}), \dots, p(E_h/c_{hg_i}), \dots, p(E_h/c_{hg_f})$$

$$4. \forall E_h \in E \exists E_h \in E : p(E_h/C) \geq p(E_h/C).$$

$$5. \forall E_h \in E : p(E_h/c_{hg_1}), \dots, p(E_h/c_{hg_i}), \dots, p(E_h/c_{hg_f}) \xrightarrow{M} \eta_M(E_h/C).$$

В даному розділі розглянуто поняття інтроформації як інтелектуальної організації людини що формує семантичну складову текстів і є джерелом його прояву в будівельних нормативах та запропоновано використання інтроформаційного методу для визначення величини несилового впливу на процес

визначення змісту. Також розроблено структуру інформаційного простору, який необхідно створити для ідентифікації змісту тексту будівельних нормативів та розроблено архітектуру системи інформаційного забезпечення процесу ідентифікації тексту будівельних нормативів.

**У третьому розділі** розроблено застосування інтроформаційного методу, який дозволяє реалізувати в штучних інтелектуальних системах рефлекторні принципи визначення змісту тексту будівельної тематики.

Впевненість в правильності визначення змісту будівельних нормативів базується на статистичних даних, таких як об'єм тексту (кількість речень), які використовуються для розрахунку впливу на зміст через інтроформаційні міри – визначеність та інформованість. Чим більша кількість експериментальних речень, тим більша довіра отриманих ймовірностям, більша точність розрахунків визначеності та інформованості і статистика отриманих результатів буде найточнішою. Це є дуже актуальним для будівельної галузі, яка містить в собі великий масив нормативних документів. Для оцінки впливу фрагментів текстів будівельних нормативів на правильність визначення ключових понять в роботі використано інтроформаційний метод. Але на відміну від класичного застосування, де розраховуються впливи на ті чи інші реакції інтелектуальної системи в дисертації вирішується задача визначення змісту будівельних нормативів по фрагментам природно мовних текстів. Для цього виконуються наступні дії:

1. Розрахунок величини впливу фрагментів текстів будівельних нормативів на ймовірність правильного визначення змісту цього тексту. Базою для визначення такого впливу є:

– ймовірність появи ключових понять в будівельних нормативах за умови, що у вхідному тексті є заданий фрагмент:

$$\forall h = \overline{1, n}: p(e_h / c_{hg}), \quad (9)$$

де  $p(e_h / c_{hg})$  – ймовірність появи ключового поняття  $e_h \in E$  за умови, що у вхідному тексті є фрагмент  $c_{hg}$ .

– безумовні ймовірності появи ключових понять в будівельних нормативах:

$$\forall h = \overline{1, n}: p(e_h), \quad (10)$$

де  $p(e_h)$  – ймовірність появи ключового поняття  $e_h \in E$ .

2. Несиловий вплив фрагментів текстів будівельних нормативів на визначення ключових понять розраховується наступним методом. Величина впливу визначається по відхиленню умовної ймовірності  $p(e_h / c_{hg})$  від безумовної:

$$\forall h = \overline{1, n}: \Delta dp(e_h / c_{hg}) = d(e_h / c_{hg}) \cdot i(e_h) - d(e_h) \cdot i(e_h / c_{hg}), \quad (11)$$

де  $\Delta dp(e_h / c_{hg})$  – додаткова визначеність появи ключового поняття  $e_h$ , яку привносить фрагмент  $c_{hg}$ ;

$d(e_h / c_{hg})$  – визначеність об'єктів і процесів, які формують появу ключового поняття  $e_h$  у випадку, коли в тексті є фрагмент  $c_{hg}$ ;

$i(e_h / c_{hg})$  – інформованість об'єктів і процесів, які формують появу

ключового поняття  $e_h$  у випадку, коли в тексті є фрагмент  $c_{hg}$ ;

$i(e_h)$  – інформованість об'єктів і процесів, які формують появу ключового поняття  $e_h$ ;

$d(e_h)$  – визначеність появи ключового поняття речення  $e_h$ .

Із формули розрахунку визначеності по ймовірностям отримаємо:

$$d(e_h / c_{hg}) = \pm 0,5 \cdot \sqrt{\frac{p(e_h / c_{hg})}{1 - p(e_h / c_{hg})} + \frac{1 - p(e_h / c_{hg})}{p(e_h / c_{hg})} - 2}; \quad (12)$$

$$d(e_h) = \pm 0,5 \cdot \sqrt{\frac{p(e_h)}{1 - p(e_h)} + \frac{1 - p(e_h)}{p(e_h)} - 2}. \quad (13)$$

Інформованість:

$$i = \sqrt{d^2 + 1}. \quad (14)$$

Тоді

$$i(e_h / c_{hg}) = \sqrt{d^2(e_h / c_{hg}) + 1}; \quad (15)$$

$$i(e_h) = \sqrt{d^2(e_h) + 1}. \quad (16)$$

3. Розрахунок загального впливу фрагментів тексту на правильність визначення ключових понять:

$$\forall g = \overline{1, n}: \Delta d_{\Sigma}(e_h) = \sum_{h=1}^m \Delta d(e_h / c_{hg}); \quad (17)$$

$$\Delta i_{\Sigma}(e_h) = \sqrt{d_{\Sigma}^2(e_h) + 1}, \quad (18)$$

де  $\Delta i_{\Sigma}(e_h)$  – сумарне доповнення інформованості про рівень правильно визначених ключових понять  $e_h$ ;

$\Delta d_{\Sigma}(e_h)$  – сумарне доповнення визначеності про рівень правильно визначених ключових понять  $e_h$ .

4. Розрахунок нового значення впливу фрагментів тексту на його зміст:

$$\forall g = \overline{1, n}: \overline{d}(e_h) = \Delta d_{\Sigma}(e_h) \cdot i(e_h) + d(e_h) \cdot \Delta i_{\Sigma}(e_h), \quad (19)$$

де  $\overline{d}(e_h)$  – нова визначеність ключових понять  $e_h$  в тексті.

5. Правило вибору:

$$\forall e_i \in E \exists e_h \overline{d}(e_h) \geq \overline{d}(e_i). \quad (20)$$

6. Розв'язком даної задачі є загальна оцінка ймовірності появи ключового поняття  $e_h$  по фрагментам текстів будівельних нормативів.

$$p(e_h) = 0,5 + \frac{\overline{d}(e_h)}{\sqrt{(\overline{d}(e_h))^2 + 1}}, \quad (21)$$

де  $p(e_h)$  – оцінка ймовірності появи ключового поняття  $e_h$ .

Із формули (8) слідує формула для оцінки ефективності наведеного методу. Така оцінка базується на розрахунку суми квадратів відхилень ймовірностей фактичного змісту:

$$E = \sum_{\text{правильно визначений фрагмент}} (p(e_h) - 1)^2 + \sum_{\text{не правильно визначений фрагмент}} (p_j(e_j))^2 \rightarrow \min, \quad (22)$$

де  $E$  – оцінка відхилення прогнозу від фактичного рівня правильно визначених ключових понять в тексті.

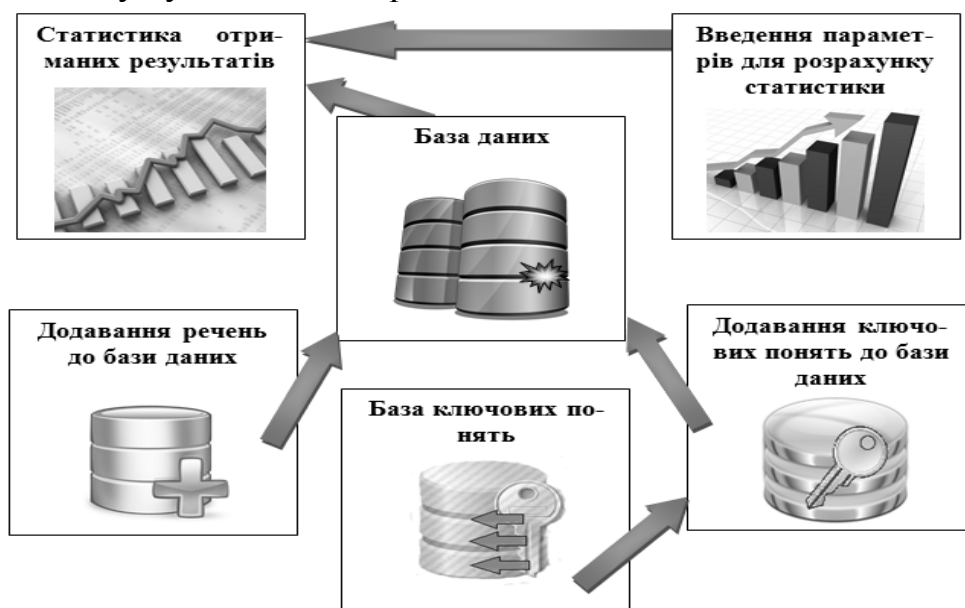
Отже запропонована модель та метод дозволяють не тільки оцінити спільну умовну імовірність появи того чи іншого змісту в тексті будівельних нормативів, але й оцінити правильність визначення такої імовірності.

**Четвертий розділ** присвячено розробці засобів інформаційної технології ідентифікації змісту текстів будівельних нормативів. Обробка експериментальних даних (результатів) інформаційної технології, розробка бази даних речень та ключових понять дозволяє проводити навчання системи на основі визначених ключових понять експертом в галузі будівництва.

В даному розділі була сформульована мета впровадження інформаційної технології ідентифікації змісту будівельних нормативів, основні принципи побудови інформаційної технології, функціональні та нефункціональні вимоги до розробки та наведені критерії якості функціонування системи. Розроблено алгоритм визначення ключових понять тексту будівельних нормативів, наведений опис вхідних даних та принципи введення їх в систему та шляхи досягнення високих результатів роботи інформаційної технології. Описано структуру бази даних, визначено сутності, атрибути сутностей, їх ідентифікатори, розроблено логічну та фізичну моделі інформаційної технології ідентифікації змісту будівельних нормативів, наведені компоненти бази даних інформаційної технології ідентифікації змісту будівельних нормативів та зв'язки між ними.

Інформаційна технологія ідентифікації змісту будівельних нормативів розроблена в середовищі MS Access з використанням мови програмування VBA Access з причини зручності застосування та багатofункціональності середовища.

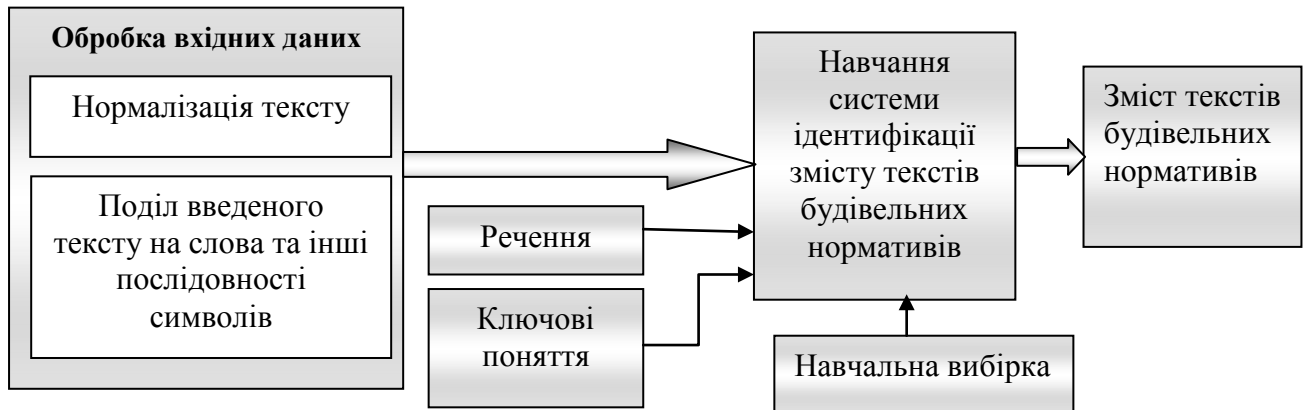
На рис. 3 представлена узагальнена структура інформаційної технології ідентифікації змісту будівельних нормативів.



**Рис. 3** Узагальнена структура інформаційної технології ідентифікації змісту будівельних нормативів

В процес обробки вхідних даних відбувається нормалізація тексту, здійснюється поділ введеного тексту на слова і інші послідовності символів. До символів належать, зокрема, знаки пунктуації та символи початку абзацу.

Схема процесу ідентифікації змісту будівельних нормативів представлена на рис. 4.



**Рис. 4** Процес ідентифікації змісту будівельних нормативів

Вхідними даними системи є речення Державних будівельних норм (ДБН) України та список ключових понять. Ключове поняття – зміст тексту будівельних нормативів. Зміст тексту будівельних нормативів представляється окремим словом або словосполученням.

При формуванні навчальної вибірки експерт може визначити декілька ключових понять для одного тексту, які не обов'язково є частиною цього тексту. Визначається саме зміст тексту, тобто про що йде мова в природно мовному тексті.

Для здійснення процесу навчання інформаційної технології ідентифікації змісту текстів будівельних нормативів була сформована навчальна вибірка, яка призначена для навчання інформаційної технології визначати ключові поняття. Програма навчається на прикладі, а після закінчення фази навчання може узагальнювати.

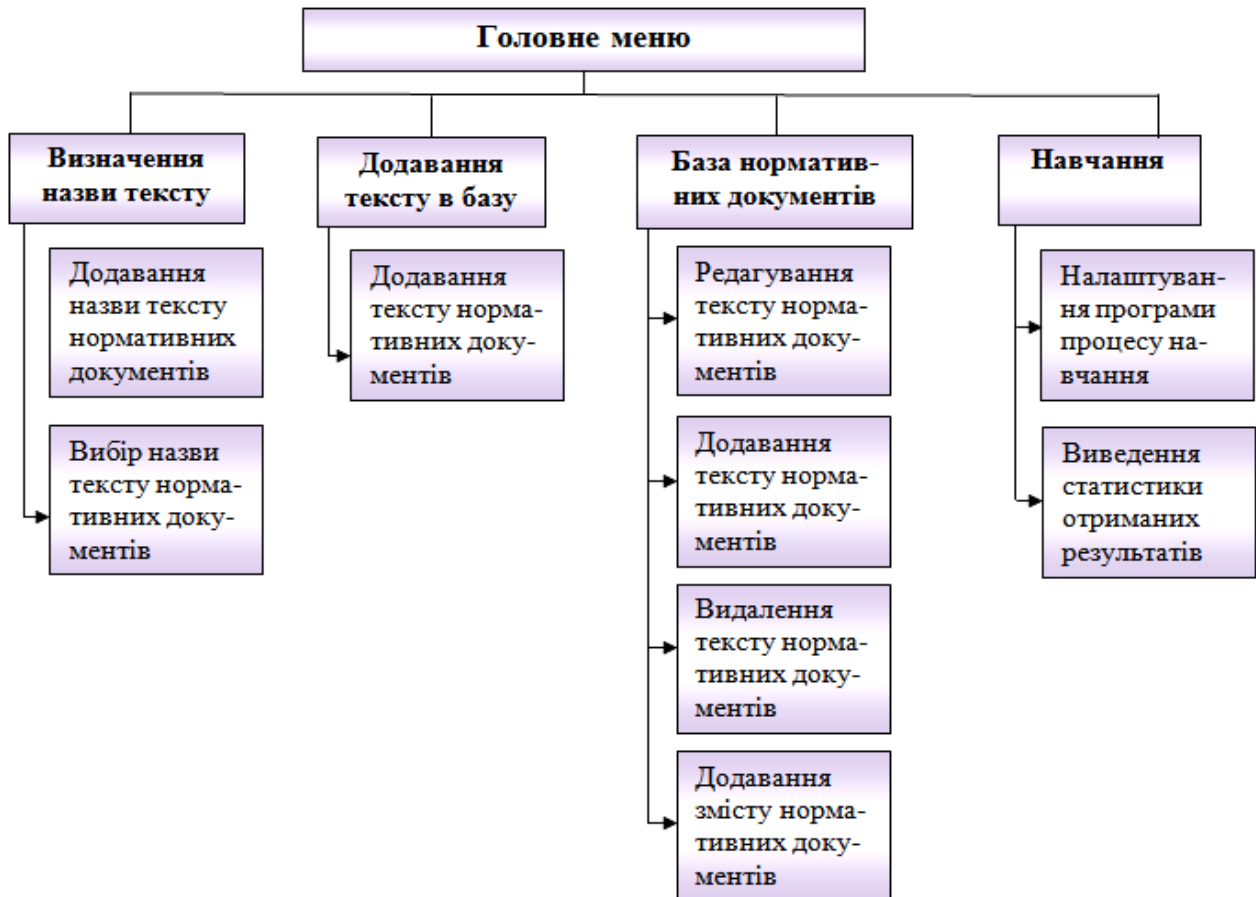
Для експериментального прикладу визначення змісту будівельних нормативів навчальна вибірка становить 50% .

Процес навчання моделі є задачею мінімізації функції вихідної похибки  $E(t)$ , де  $t$  – номер ітерації навчання. В процесі навчання похибка навчання монотонно зменшується.

Технологія не просто обробляє і зберігає поданий приклад, а й визначає закономірності в даних для навчання. При певному наборі параметрів модель повинна відтворювати закономірності, отримані під час навчання, для подальшого прогнозування.

Меню інформаційної технології ідентифікації змісту будівельних нормативів складається з вікна визначення назви тексту, додавання тексту в базу, бази нормативних документів та навчання.

Структура меню інформаційної технології ідентифікації змісту будівельних нормативів наведена на рис 5.



*Рис. 5* Структура меню інформаційної технології ідентифікації змісту будівельних нормативів

В розділі наведено приклад ідентифікації змісту текстів будівельних нормативів розробленими засобами та виконано порівняння з іншими програмними засобами (TextAnalys). Результати показані засобами «Інформаційна технологія ідентифікації змісту будівельних нормативів» (87% правильної ідентифікації) набагато кращі від тих, які показала програма TextAnalys 25%).

З викладеного матеріалу можна зробити висновок, що на сьогоднішній день немає такої технології, яка б ефективно вирішувала задачі лінгвістичної обробки текстів будівельних нормативів. Тому приклад розробки такої технології, наведений в дисертаційній роботі, яка показала високі результати ідентифікації змісту текстів будівельних нормативів, може стати основою для подальших досліджень в цьому напрямку, які сумістили б в собі засоби штучного інтелекту, моделювання та прогнозування виконання будівельних проектів.

Зазначені результати дослідження знайшли практичне застосування в роботі ряду промислових будівельних підприємств і пройшли належну наукову апробацію на семінарах і конференціях.

Отже, інформаційна технологія ідентифікації змісту текстів будівельних нормативів на основі рефлєкторного алгоритму має наступні переваги:

- здатність до вирішення широкого кола завдань в деякій неформалізованій проблемній області будівництва;

- здатність витягати знання з даних і представляти їх у вигляді формалізованих моделей знань;
- моделювання механізмів інтелектуальної діяльності людини;
- можливість інтегрувати інформаційну технологію та застосувати рефлексорні методи в будь-якій галузі та виробництві.

## **ВИСНОВКИ**

Робота присвячена вирішенню актуальної науково-технічної задачі розробки методів та моделей ідентифікації змісту будівельних нормативів. В роботі були отримані такі наукові та практичні результати:

1. Проведено аналіз методів і моделей управління інформацією в нормативній базі будівництва шляхом наукових досліджень.

2. Виконано аналіз побудови архітектури інструментів ідентифікації змісту природномовних текстів, що дало можливість врахувати специфіку будівельної галузі.

3. Започатковано науково-методологічний та понятійний базис управління процесом лінгвістичного аналізу текстів будівельних нормативів. Виконано класифікацію задач в сфері лінгвістичного аналізу природномовних текстів, для вирішення яких необхідне застосування науково-методичних засобів штучного інтелекту в спеціалізованих інформаційно-аналітичних системах, які забезпечують зменшення часу на пошук необхідної інформації в будівельних нормативах.

4. Досліджені реакції зовнішніх впливів та впливу людського мислення на хід ідентифікації змісту природномовних текстів будівельних нормативів.

5. Побудовано модель формалізації змісту інформаційного масиву будівельних нормативів.

6. Запропоновано використання рефлексорного підходу до побудови інформаційної технології ідентифікації змісту будівельних нормативів для проектно-будівельних підприємств.

7. Розроблено практичні інструменти ідентифікації змісту текстів будівельних нормативів.

8. Розвинуто методи та моделі побудови рефлексорних інтелектуальних систем, які базуються на концепції несилової взаємодії в області інтелектуальної діяльності людини і описуються математичним апаратом теорії несилової взаємодії.

9. Наукові положення, висновки, пропозиції і рекомендації дисертаційної роботи можуть бути використані для практичної організації ведення нормативної бази в будівництві. Робота впроваджена в проектно-будівельних компаніях України та використовується в навчальному процесі в Київському національному університеті будівництва і архітектури, так як є універсальним засобом для обробки будь-яких текстів необмежених об'ємів.



## СПИСОК ОПУБЛІКОВАНИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

### *Праці у закордонних виданнях:*

1. Шабала Е.Е. Интеллектуальная система распознавания текста строительной тематики [Текст] / Е.Е. Шабала, Ю.Н. Тесля // Международный научный журнал «Наука и мир», Волгоград. – 2014 г. – Вып. 7 (11) . – с. 25-28.

*Автором описано експериментальний приклад роботи інтелектуальної системи розпізнавання текстів будівельної тематики.*

### *Праці у фахових виданнях:*

2. Задоров В.Б. Використання імітаційного моделювання для формування нормативів витрат ресурсів для будівельних процесів [Текст] / В.Б. Задоров, В.Т. Шпирний, Є.Є. Шабала // Управління розвитком складних систем, КНУБА. – 2013. – Вип. 13. – с. 97-103.

*Автором запропоновано підхід до використання імітаційного моделювання для формування нормативів витрат ресурсів для будівельних процесів.*

3. Шабала Є.Є. Застосування нечіткої логіки у формуванні моделі технологічної карти на обладнання перегородок комплектної системи КНАУФ [Текст] / Є.Є. Шабала, В.В. Ключова // Управління розвитком складних систем. – 2013р. – Вип. 17. – с. 100-106.

*Автором запропонована модель використання нормоутворюючих факторів для технологічної карти будівельних процесів.*

4. Тесля Ю.М. Рефлекторна система формалізації змісту довільних текстів будівельної тематики [Текст] / Ю.М. Тесля, Є.Є. Шабала // Управління розвитком складних систем. – 2013р. – Вип. 16. – с. 119-123.

*Автором запропоновано використання інтроформаційного методу для визначення ключових понять текстів будівельних нормативів, який забезпечує розрахунок несилового впливу фрагментів речень на вибір ключових понять.*

5. Терентьев О.О. Основы организации нечеткого вывода для задачи диагностики технического stanu будівель [Текст] / О.О. Терентьев, Є.Є. Шабала, Б.С. Малина // Управління розвитком складних систем. – 2015р. – Вип. 22. – с. 138-143.

*Автором описаний аналіз нечітких систем, які дозволяють вирішувати задачі створення інформаційної технології для підтримки прийняття рішень щодо діагностики технічного стану будівель.*

6. Тесля Ю.М. Особливості обробки природномовного тексту будівельної тематики [Текст] / Ю.М. Тесля, Є.Є. Шабала // Вісник Хмельницького національного університету. Технічні науки, 2015. – № 1. – с. 172-174.

*Автором запропоновано для розрахунку несилового впливу фрагментів тексту на вибір ключових понять в нормативній базі будівельної галузі застосувати інтроформаційний підхід.*

### *Матеріали міжнародних наукових конференцій:*

7. Шабала Є.Є. Створення сучасної структури інформаційної системи нормативної бази будівництва на основі врахування нормоутворюючих факторів

викоритання ресурсів різних типів [Текст] / Є.Є. Шабала // Тези доповідей наукової конференції молодих вчених, аспірантів і студентів – К.: КНУБА, 2011. – с. 162-163.

8. Шабала Є.Є. Моделювання ресурсів на одиничному комплексі елементарних будівельних операцій [Текст] / Є.Є. Шабала // Тези доповідей наукової конференції молодих вчених, аспірантів і студентів. – К.: КНУБА, 2012. – с. 170-171.

9. Шабала Є.Є. Розробка моделей і засобів аналізу та синтезу інформаційної системи нормативної бази ресурсів в підготовці будівництва [Текст ] / Є.Є. Шабала // Тези доповідей всеукраїнської науково-практичної Інтернет-конференції «Автоматизація та комп'ютерно-інтегровані технології у виробництві та освіті: стан, досягнення, перспективи розвитку». – Черкаси.: ЧНУ ім. Б. Хмельницького, ЧІБС, ЧДУ ім. П. Могили, 2013. – с. 30-34.

10. Шабала Є.Є. Інформаційні інтелектуальні системи для ведення нормативної бази ресурсів в будівельній сфері [Текст] / Є.Є. Шабала // Тези доповідей всеукраїнської науково-практичної Інтернет-конференції «Автоматизація та комп'ютерно-інтегровані технології у виробництві та освіті: стан, досягнення, перспективи розвитку». – Черкаси.: ЧНУ ім. Б. Хмельницького, ЧІБС, ЧДУ ім. П. Могили, 2014. – с. 12-13.

11. Тесля Ю.М. Ідентифікація змісту текстів методами штучного інтелекту [Текст] / Ю.М. Тесля, Є.Є. Шабала // Тези доповідей першої міжнародної науково-практичної конференції «Управління розвитком технологій». Тема: Управління програмами та проектами в умовах глобальної фінансової кризи. – К.: КНУБА, 2014. – с. 93-94.

12. Тесля Ю.М. Лінгвістична обробка текстів будівельних нормативів [Текст] / Ю.М. Тесля, Є.Є. Шабала // Тези доповідей другої міжнародної науково-практичної конференції «Управління розвитком технологій». Тема: Технологія управління інформацією та знаннями в сучасному суспільстві.– К. : КНУБА, 2015. – с. 113-114.

## АНОТАЦІЯ

**Шабала Є.Є. Інформаційна технологія ідентифікації змісту будівельних нормативів.** – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.06 – інформаційні технології. – Київський національний університет будівництва і архітектури, м. Київ, 2015 р.

Дисертація присвячена задачі розробки інформаційної технології ідентифікації змісту будівельних нормативів. Розроблена інтроформаційна модель аналізу вихідного тексту нормативів для виділення семантичної інформації, яка міститься в тексті. Запропоновано нову міру оцінки впевненості у правильності реагування на зміст тексту нормативної бази через оперування сумісними умовними ймовірностями. Запропоновано нові визначення ключових понять тексту на основі рефлєкторно-імовірнісного підходу, що характеризується описом поведінки системи на основі рефлєксів. Удосконалено математичну модель рефлєкторного оперування інформацією в природних і штучних інтелектуальних системах для визначення сумарного несилового впливу та адекватної реакції на цей вплив. Показано, що розроблена інформаційна технологія ідентифікації змісту будівельних нормативів

може застосовуватися в діяльності будівельних компаній. Наведено приклади її застосування.

**Ключові слова:** зміст будівельних нормативів, ідентифікація змісту будівельних нормативів, рефлекторно-ймовірнісний підхід, реакція на вплив.

## АННОТАЦІЯ

**Шабала Е.Е. Информационная технология идентификации смысла строительных нормативов. - Рукопись.**

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.06 - Информационные технологии. - Киевский национальный университет строительства и архитектуры, г. Киев, 2015 г.

Диссертация посвящена решению задачи разработки информационной технологии идентификации смысла строительных нормативов. Были выявлены основные принципы развития нормативного обеспечения строительства и проведен анализ существующих подходов к разработке интеллектуальных систем. Выявлено, что построение информационной технологии идентификации смысла является одной из составляющих успешного управления нормативной базой и подготовки процессов строительства.

Показано, что основное внимание при проектировании информационной технологии идентификации смысла строительных нормативов следует уделять ее способности интеллектуальной обработки данных, через обучение и обеспечения возможности работы со знаниями на основе формализации текста.

Разработана интроформационная модель анализа исходного текста нормативов для выделения семантической информации, которая содержится в тексте. Предложена новая мера оценки уверенности в правильности реагирования на смысл текста нормативной базы через оперирование совместимыми условными вероятностями по частным условным вероятностям. Предложены новые определения ключевых понятий текста на основе рефлекторно-вероятностного подхода, характеризующегося описанием поведения системы на основе рефлексов. Усовершенствована математическая модель рефлекторного оперирования информацией в естественных и искусственных интеллектуальных системах для определения суммарного несилового воздействия и адекватной реакции на это воздействие. Автором рассмотрены процессы анализа лингвистической обработки текстов, этапы оценки вероятности правильно определенного смысла текстов нормативных документов. Для обеспечения расчета несилового влияния фрагментов предложений на выбор ключевых понятий для нормативной базы в строительной отрасли предложено применение интроформационного метода.

Сформулированы основные принципы построения информационной технологии, функциональные требования к разработке и приведены критерии качества функционирования системы. Показано, что разработанная информационная технология идентификации смысла строительных нормативов может применяться в деятельности строительных компаний. Приведены примеры ее использования. Разработанная информационная технология идентификации смысла нормативных документов внедрена в деятельность проектно-строительных

предприятий Украины.

**Ключевые слова:** смысл строительных нормативов, идентификация смысла строительных нормативов, рефлекторно-вероятностный подход, реакция на воздействие.

## ABSTRACT

**Shabala Y.Y. Information technology identification the meaning of building standards.** - Manuscript.

Thesis for scientific degree of candidate of technical sciences, specialty 05.13.06 - Information technologies. - Kyiv National University of Construction and Architecture, Kyiv, 2015.

Dissertation is devoted to the development of information technology, identification the meaning of building standards based on the methods of artificial intelligence and mathematical apparatus of the theory of non-force interaction. As part of the technology developed information model analysis of the input text to highlight the standards of semantic information that is contained in the text. Also proposed a new measure of confidence in the correctness of the assessment of response to the contents of the proposal to the regulatory framework through consistent handling conditional probability of occurrence of text fragments on individual conditional probabilities. This assessment will allow to build intelligent software system is not based on heuristics and using the mathematical apparatus of the theory of non-force interaction. Also, proposed new definitions of the key concepts of the text on the basis of reflex-probabilistic approach, which is determined by the description of the system behavior based on reflexes. The mathematical model of the reflex operating information in the natural and artificial intelligent systems for the determination of total non-violent interaction and adequate responses to this impact informational technologies. The information technology of identification the meaning of the texts of construction standards that can be applied in the activities of construction companies.

**Keywords:** system identification word meaning building norm of Islands, the theory of non-force interaction, assessment of the degree of confidence, reflex-probabilistic approach, the reaction to the impact.