

УДК 004.056

[0000-0002-5902-8628] **Я. В. Тарасенко**, канд. техн. наук,
e-mail: yaroslav.tarasenko93@gmail.com
[0000-0003-2039-2841] **В. Г. Бабенко**, д-р техн. наук, доцент
e-mail: v.babenko@chdtu.edu.ua
Черкаський державний технологічний університет
б-р Шевченка, 460, м. Черкаси, 18006, Україна

МЕТОД КОМП'ЮТЕРИЗОВАНОГО МОДИФІКУВАННЯ АНГЛОМОВНОГО ТЕКСТУ НА ОСНОВІ ПСИХОСЕМАНТИЧНИХ ВЛАСТИВОСТЕЙ

У статті досліджено можливість використання психосемантичних властивостей з метою захисту текстової інформації з обмеженим доступом. Розроблений метод комп'ютеризованого модифікування англomовного тексту дозволяє підвищити ефективність запобігання незаконному розповсюдженню та використанню текстової інформації порівняно з методами, які не враховують психосемантичну складову. Це дає змогу захистити модифіковані семантично неподільні елементи від можливості їх спотворення чи видалення, враховуючи загрозу потенційних інформаційних впливів. Проведено моделювання процесу автоматизованого модифікування англomовних текстів за допомогою комп'ютерних засобів. Результати порівняння з аналогічними підходами дозволяють зробити попередні висновки про вищу ефективність під час роботи з текстами середнього та високого об'єму. Отримані експериментальні дані також дають можливість визначити стійкість модифікованих за допомогою розробленого методу текстів до застосування спеціалізованих засобів семантичного видалення наявних маркерів.

Ключові слова: комп'ютеризована модифікація тексту, психосемантичні властивості, семантично неподільна складова, інформація з обмеженим доступом, квантово-семантичний аналіз, суб'єктивне сприйняття семантики.

Вступ. Сьогодні країна стикається з викликами та загрозами інформаційній безпеці, вирішення яких вкрай необхідне як для держави, так і для інших країн-партнерів. Згідно зі Стратегією інформаційної безпеки [1], з одного боку, існування ефективної системи захисту і протидії нанесенню шкоди через поширення негативних інформаційних впливів, до яких відносяться деструктивна пропаганда, несанкціоноване розповсюдження, використання та порушення цілісності інформації з обмеженим доступом, є необхідною умовою інформаційної безпеки України. З другого боку, аналіз загроз та викликів інформаційній безпеці доводить, що застосування технологій інформаційного протиборства проти України також має вплив і на інші держави, а отже, як висновок, розгляд саме англomовного тексту дозволить нівелювати шкідливі наслідки як проти України, так і проти інших держав.

Інсайдерські атаки на інформацію з обмеженим доступом, зокрема на текстову інформацію, спричинили втрати в середньому \$11,45 млн. за 12 місяців станом на 2020 р. за даними незалежного дослідження Ponemon

Institute [2]. Звідси випливає необхідність розробки комп'ютерних інформаційних систем, здатних швидко та ефективно протидіяти подібним загрозам. Існує необхідність вдосконалення систем електронно-цифрового підпису та базових методів, що лежать в основі методології створення DLP (Data Loss Prevention) систем. Зокрема, задача накладання цифрових водяних знаків (ЦВЗ) на текстову інформацію потребує вдосконалення для підвищення стійкості подібних міток. Психосемантичні методи дозволяють підвищити надійність вбудовування ЦВЗ у текст шляхом його модифікації. Звідси випливає задача створення методу комп'ютеризованого модифікування англomовного тексту на основі психосемантичних властивостей, який зможе підвищити рівень захисту інформації з обмеженим доступом і може бути використаний в інформаційному протиборстві для забезпечення інформаційної безпеки держави.

Аналіз літератури показав наявність значної кількості методів генерації текстів [3-5], однак недостатню кількість методів модифікації вже існуючих текстів [6-7], що зумов-

лює потребу в подібних методах. Крім того, базові підходи, застосовані в існуючих методах генерації тексту хоч і можуть бути використані для вирішення окремих завдань модифікації тексту, однак повною мірою не можуть задовольнити усі потреби в методологічному апараті. В той же час існуючі методи модифікації не враховують психосемантичних особливостей автора тексту, що унеможливає їх використання в поточному контексті. При цьому в розробці методу, якому присвячена стаття, корисним буде використання методу розрахунку когерентності [8] і попередніх розробок за основною тематикою [9-12] як методологічного базису для вирішення поставленої задачі, адже необхідність урахування аспекту модифікації тексту разом із генерацією також зумовлює актуальність дослідження.

Мета дослідження полягає у підвищенні ефективності запобігання незаконному розповсюдженню та використанню інформації з обмеженим доступом шляхом вдосконалення методологічної основи, враховуючи протидію інформаційним впливам.

В роботі ставляться наступні **задачі дослідження**:

1. Визначення імовірності успішного модифікування семантично неподільної складової.

2. Модифікація семантично неподільних складових за психосемантичними властивостями.

3. Забезпечення контекстуальної єдності модифікованої семантично неподільної складової.

Для визначення можливості модифікування неподільної семантичної складової, отриманої завдяки методу її виявлення та збереження [11], пропонується використання теорії конгруентності Озгуда і Танненбаума [13] із застосуванням методу додавання семантичного шуму [12]. Для модифікації тексту використовується психолінгвістичний профіль, сформований за методом квантово-семантичного психолінгвістичного аналізу англомовного тексту [10]. Метод комп'ютерного формування англомовного тексту відповідно до психолінгвістичного профілю [9] виступає основою для проведення маніпуляцій із семантикою тексту. З метою забезпечення контекстуальної єдності модифікованих частин, вдосконалюється метод розрахунку конгруентності тексту [8], для чого застосовується визна-

чення семантичного диференціалу із вдосконаленим в [14] методом факторіального аналізу. Аналіз надійності здійснювався із застосуванням методу семантичного стиснення текстової інформації [15].

Виклад основного матеріалу. Виходячи з того, що в дослідженні отримали подальший розвиток раніше розроблені методи і вже отримані результати члена авторського колективу статті, доцільно надати базові визначення для правильного розуміння використаних понять.

Темою в цьому випадку виступає структурно-семантичний функціональний блок, виражений лексико-семантичними засобами, що несе в собі відбиток відомої інформації.

Ремою виступає блок, який несе в собі відбиток невідомої інформації.

Диремою виступає функціональний блок, в якому частина інформації є вже відомою, а інша частина ще невідомою.

Психосемантичні властивості характеризують індивідуальне семантичне сприйняття та ставлення суб'єкта і базуються на його категоріальному апараті сприйняття об'єктивної дійсності.

Конгруентність передбачає встановлення балансу ставлення суб'єкта до об'єкта мовлення.

Семантичним шумом можна назвати ті смислові елементи, які не впливають на об'єктивне сприйняття основної теми та підтем тексту.

Світ сприйняття виступає окремою властивістю для аналізу семантичних структур з погляду їх можливого сприйняття суб'єктом.

Визначення можливості модифікування семантично неподільної складової. Базуючись на даних, отриманих в [11], використовуючи комп'ютерну інформаційну систему, можна визначити тему та рему англомовного тексту, а сума рем та дирем виступає неподільною семантичною складовою. Модифікація саме неподільної семантичної складової дозволить забезпечити цілісність даних та підвищити ефективність процедури модифікації тексту в напрямі захисту модифікованих семантичних елементів від видалення чи повторної зміни, що спричинить спотворення маркера. Крім цього, захист маркера від спотворення в результаті можливого відтворення тексту зловмисником по пам'яті можуть забезпечити лише семантичні одиниці, а ні в якому разі не лексичні, які легко замінити без втрати основного смислового навантаження тексту.

Виходячи з результатів дослідження [11] та виведеної формули для визначення множини збережених рем R^* , можна отримати суму рем та дирем, яка становить неподільну семантичну складову, враховуючи маркер психолінгвістичного портрету зловмисника. Однак цього недостатньо для подальшої модифікації тексту у зв'язку з необхідністю визначення можливості модифікації визначеної неподільної семантичної складової шляхом обрахунку імовірності успішної модифікації на основі психосемантичних властивостей зловмисника. Базуючись на теорії конгруентності Озгуда і Танненбаума [13] і результатах застосування теорії в методі додавання семантичного шуму за індивідуальною семантичною лінією пропагандиста [12], можна адап-

тувати методику для визначення імовірності успішного модифікування неподільної семантичної складової англomовного тексту. Перш за все, обраховується негентропія неподільної семантичної частки:

$$H(R^*) = - \sum_{i=1}^m P(R^*) \cdot \log_2 \frac{1}{P(R^*)} . \quad (1)$$

де R^* – неподільна семантична складова.

Негентропія є необхідною умовою визначення рівня впорядкованості неподільних семантичних складових.

Таким чином, повна імовірність події, що неподільну семантичну складову можна модифікувати, обчислюється за формулою

$$P(R_{\text{mod}}^*) = \sum_{i=1}^n P([C \text{ to } H(R^i)] \rightarrow 0) \cdot P(R_{\text{mod}}^* | [C \text{ to } H(R^i)] \rightarrow 0), \quad (2)$$

де $C \text{ to } H(R^i)$ – зміна ставлення особи, психосемантичні властивості якої використовуються до неподільної семантичної складової. В ідеальному випадку зміна ставлення повинна дорівнювати нулю, однак обираються лише ті неподільні семантичні складові, які максимально наближені до нуля. При цьому зміна ставлення до неподільної семантичної складової може бути обчислена за рівнянням

$$C \text{ to } H(R^*) = \begin{cases} C \text{ to } H(R^*) = \frac{H(R^*)}{H(R^*) + |H(R_{\text{mod}}^*)|} \cdot p' \\ C \text{ to } H(R_{\text{mod}}^*) = \frac{|H(R_{\text{mod}}^*)|}{H(R^*) + |H(R_{\text{mod}}^*)|} \cdot p' \end{cases} \quad (3)$$

де $C \text{ to } H(R_{\text{mod}}^*)$ – зміна ставлення до модифікованої неподільної семантичної складової, p' – величина сумарної зміни ставлення.

Процедура комп'ютерного модифікування англomовного тексту. Визначивши потенційну можливість модифікації семантично неподільної складової тексту, необхідно провести зміни на рівні сприйняття. При цьому, за аналогією з односторонньою функцією, потенційно неможливо дізнатись, яка саме семантична складова була модифікована, навіть використавши той самий підхід визначення теми-реми та потенційних одиниць для модифікації. Це досягається за рахунок використання стеганографічного протоколу вбудовування ЦВЗ у текст на основі приватного

ключа, яким і виступає сформований психолінгвістичний портрет конкретної особи (в цьому контексті зловмисника). Для вирішення поставлених у статті задач не є необхідною розробка стеганографічного протоколу. Наразі важливо сформулювати методологічний апарат для його подальшого вбудовування в готовий текст, на що і розрахований пропонований метод модифікування.

Модифікування тексту шляхом модифікації його семантично неподільних складових можливе, спираючись на психосемантичні властивості, зокрема на категоріальний апарат зловмисника та його психолінгвістичний профіль, визначений за допомогою методу квантово-семантичного психолінгвістичного аналізу англomовного тексту [10] із використанням методу комп'ютерного формування англomовного тексту відповідно до психолінгвістичного профілю [9]. Водночас адаптація методу [11] до квантово-семантичного аналізу дозволяє значно підвищити синергію цих методів і, як результат, підвищити ефективність модифікування тексту.

Таким чином, за формулою (4) можна сформувати семантично однорідний необхідний фрагмент тексту, яким варто замінити існуючу в тексті неподільну семантичну складову:

$$R_{\text{mod}} = \lambda \hat{P}R^* [H(R^*)(\hat{P}R^*)(f(x')dx)], \quad (4)$$

де \hat{P} – оператор перестановки, $f(x')dx$ – функція, що описує суб'єктивну семантичну лі-

нію, λ – множник, який відрізняє сприйняття модифікованої семантичної лінії від суб'єктивної.

В свою чергу, формула (5) описує процес модифікування англomовного тексту на основі психосемантичних властивостей:

$$T_{\text{mod}} = \frac{\prod_{k=1}^n (R^*)_k}{\prod_{k=1}^m (R_{\text{mod}}^{S'_{1_n}})_k} + \prod_{k=1}^m (R_{\text{mod}}^{S'_{1_n}})_k, \quad (5)$$

де S'_{1_n} – множина синонімічних лексико-семантичних конструкцій, характерних для психолінгвістичного профілю та розподілених на e елементів, які відповідають категоріальним одиницям демографічного профілю.

Забезпечення контекстуальної єдності семантично неподільної складової, враховуючи психосемантичні властивості. Зміни найбільш вдалих елементів недостатньо для повноцінного модифікування тексту в комплексі, оскільки зростає ризик виявлення модифікованого фрагмента. Для підвищення ефективності та надійності процедури модифікації необхідно врахувати контекстуальну залежність і звернути увагу на психосемантичні властивості автора. Це дозволяє зробити модифікований під задачі дослідження метод розрахунку когерентності тексту [8]. Щоб вдосконалити метод, необхідно враховувати

$$\frac{1}{m(m-1)} \sum_{i,j=1}^m (\overline{W}'_i - \overline{T}_{\text{mod}})_j^2 = \frac{2m^2}{m(m-1)} \frac{\left[\sum_{i=1}^{|R^*|} P(R_{\text{mod}}^*) \cdot \log_2 \frac{1}{P(R_{\text{mod}}^*)} \right]}{\prod_{k=1}^m (R_{\text{mod}}^{S'_{1_n}})_k}, \quad (7)$$

де W' – вектор світів сприйняття тексту.

В іншому послідовність дій для визначення семантичного диференціалу не відрізняється від класичного алгоритму дій, модифікованого в [14]. Також наявна необхідність визначення світів сприйняття.

Процедуру оцінювання задачі вставки, описаної у [8], доцільно використовувати і для оцінювання запропонованого в статті методу. Універсальність підходу досягається завдяки адаптації методологічного апарату вдосконалюваного методу до застосування до попередніх досліджень за темою статті. Це дає змогу з максимально можливою точністю використовувати результати експериментального дослідження в роботі [8] з метою порів-

няння з попередніми експериментальними результатами моделювання роботи методу, отриманими при дослідженні запропонованого методу комп'ютеризованого модифікування англomовного тексту.

психосемантичні властивості шляхом визначення семантичного диференціалу, в основі процесу визначення якого лежить вдосконалений в [14] метод факторіального аналізу для урахування світів сприйняття семантичної частки. А той факт, що вдосконалений пошук ортогональних проєкцій дозволяє знайти і довжину вектора семантики, то представлення модифікованої семантично неподільної складової тексту у векторному вигляді дозволить зменшити кількість аспектів методу [8], які необхідно вдосконалити, та підвищити ефективність його роботи.

Таким чином, вектор модифікованого тексту набуде вигляду

$$\overline{T}_{\text{mod}} = \frac{1}{|R^*|} \sum_{k=1}^{|R^*|} \overline{(R_{\text{mod}})_k} \cdot \overline{V}, \quad (6)$$

де \overline{V} – вектор класифікації, координати якого корелюють з усіма неподільними семантичними складовими, що співвідносяться з категоріальним апаратом психолінгвістичного портрету, $|R^*|$ – потужність множини усіх неподільних семантичних складових.

Пошук ортогональних проєкцій буде відбуватися за модифікованою під задачі цього дослідження формулою наступного вигляду:

няння з попередніми експериментальними результатами моделювання роботи методу, отриманими при дослідженні запропонованого методу комп'ютеризованого модифікування англomовного тексту.

Результати досліджень. В роботі розроблено метод комп'ютеризованого модифікування англomовного тексту, що базується на психосемантичних властивостях для впровадження ЦВЗ у текст за умови його захисту від модифікації чи видалення, що дає змогу підвищити ефективність запобігання незаконному розповсюдженню та використанню інформації з обмеженим доступом шляхом впровадження маркера у семантично неподільну складову.

При цьому було вдосконалено метод розрахунку конгруентності тексту для врахування семантичних властивостей шляхом застосування модифікованого факторіального аналізу на основі визначення світів сприйняття семантичної частки.

Визначення можливості модифікування семантично неподільної складової відбувається шляхом обрахунку імовірності успішної модифікації на основі психосемантичних властивостей за рахунок застосування теорії конгруентності та результатів аналізу методу додавання семантичного шуму, що дало змогу визначити імовірність успішного модифікування тексту.

Модифікація англомовного тексту відбувається шляхом модифікації семантично неподільних складових за психосемантичними властивостями, які базуються на категоріальному апараті та психолінгвістичному профілі, що дозволило більш ефективно вбудовувати маркер у семантично неподільну складову тексту для його захисту.

Забезпечення контекстуальної єдності відбувається на основі модифікованого під задачі дослідження методу розрахунку коге-

рентності тексту на основі визначення семантичного диференціалу за рахунок вдосконаленого методу факторіального аналізу, що дозволило врахувати світи сприйняття семантичної частки і підвищити загальний рівень ефективності модифікації тексту та природності його сприйняття з метою захисту модифікованих елементів.

Для перевірки надійності та ефективності методу було обрано тестову групу осмислених текстів публіцистичного стилю об'ємом від 100 до 50 000 символів загальною кількістю 70 текстів. Вбудовування відбувалося в автоматизованому вигляді: комп'ютеризоване модифікування з мануальним формуванням ЦВЗ.

До того ж, не варто залишати без уваги надійність методу в умовах інформаційного протиборства, окрім ефективності модифікації тексту.

Таким чином, застосувавши найбільш надійний спосіб знищення надлишкової семантичної інформації – семантичне стиснення тексту за допомогою методу [15], можна визначити надійність та захищеність запропонованого методу (рисунок 1).

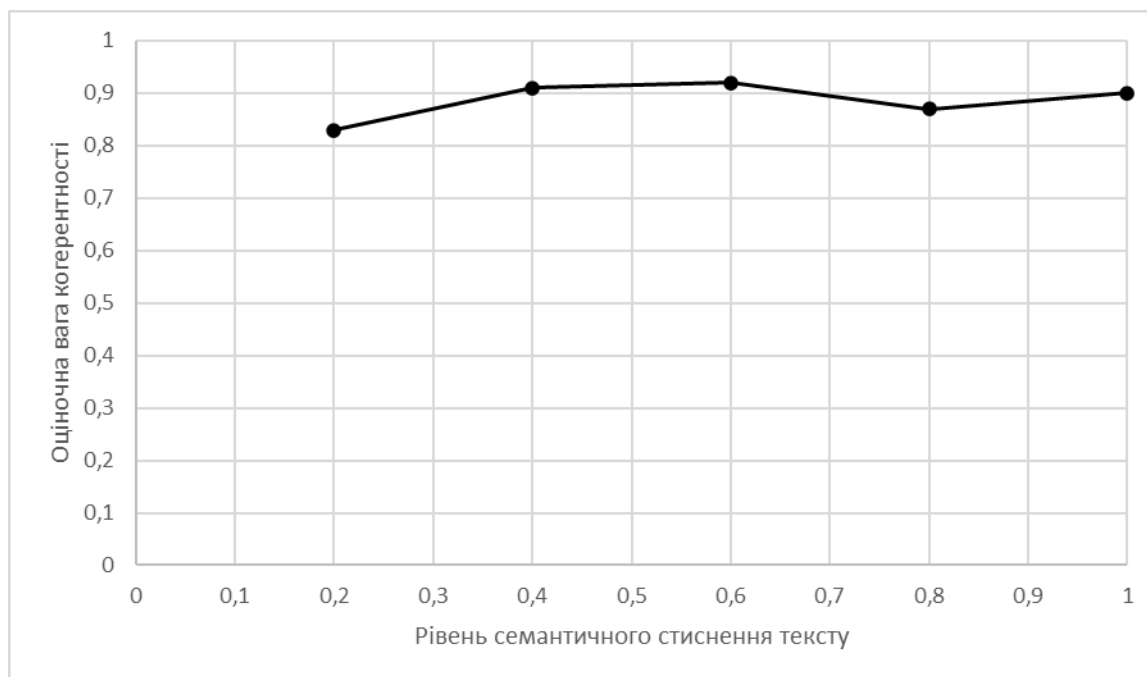


Рисунок 1 – Зміна когерентності при семантичному стисненні

Таким чином, текст стискався до значень, які в окремих випадках досягали 20% від початкового об'єму тексту. При цьому перевірялась семантична схожість модифіко-

ваних об'єктів з початковими на основі когерентності. І хоча виявлено, що при порогових значеннях у 20% схожість досягала 83% і була величиною нестабільною, на яку впливало

багато сторонніх змінних, однак загальна схожість не була нижчою за 90%. Окрім того, виявлено, що стиснення не видалило жодного модифікованого об'єкта, а це, в свою чергу, доводить ефективність приховування ЦВЗ в семантично неподільних складових, які видалити неможливо через наявність межі семантичного стиснення.

Обговорення результатів. В роботі [8] зазначаються найкращі експериментальні дані у моделі DM (distributed memory) підходу MSV (multiple similar vertex). Наведені експериментальні дані можна використати як аналог для порівняння та аналізу функціонування запропонованого в статті методу. Як видно з рисунка 2, найкращу оціночну вагу аналогічний підхід показує лише за найниж-

чого порогового значення (штрихова лінія на графіку). І хоча недоліком є той факт, що результати запропонованого методу нижчі за умови мінімального порогового значення (суцільна лінія на графіку), однак зі зростанням порогового значення зростає і ефективність методу. Це пояснюється впливом порогового значення на загальний об'єм тексту, а як відомо, зі збільшенням об'єму збільшується кількість імовірних семантично неподільних складових, які, за оцінкою, можуть бути використані для модифікації тексту за психосемантичними властивостями. При цьому для аналогічних методів збільшення об'єму означає зменшення імовірності позитивної кореляції доданого фрагмента з контекстом.

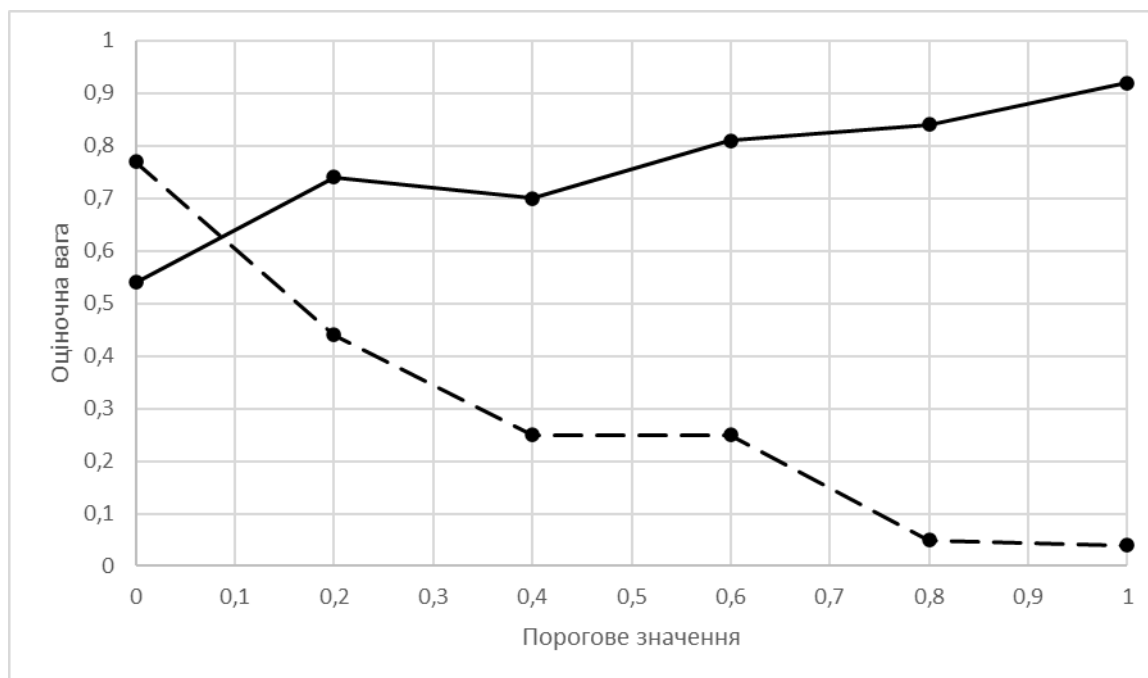


Рисунок 2 – Залежність оціночної якості модифікації тексту від порогового значення

Можна припустити, що суттєва різниця значень зумовлена різницею в підходах до аналізу української та англійської мови, застосуванні квантово-семантичного аналізу та врахуванні психосемантичних властивостей. На жаль, не відомі певні деталі проведеного у [8] експериментального дослідження, зокрема дискурсні характеристики аналізованих текстів, що не дозволяє зі стовідсотковою точністю зіставити досліджувані дані. Однак основні вимоги до експерименту були витримані, а досліджувані характеристики дозволяють мінімізувати розбіжність та забезпечити максимально можливу об'єктивність результатів.

Висновки. Наукова новизна полягає в тому, що розроблений метод комп'ютеризованого модифікування англійського тексту дозволяє підвищити ефективність запобігання незаконному розповсюдженню та використанню інформації з обмеженим доступом шляхом впровадження маркера у семантично неподільну складову тексту та модифікації її на основі психосемантичних властивостей, що дозволило розробити методологічний апарат для впровадження ЦВЗ у текст за умови його захисту від модифікації чи видалення та врахувати протидію інформаційним впливам завдяки вдосконаленому методу розрахунку кон-

груєнтності тексту для врахування психолінгвістичного портрету зловмисника.

Метод дає змогу одночасного вирішення двох задач: захисту текстових документів та здійснення інформаційного протидиверсія за рахунок формування універсального методологічного апарату для протидії кібердиверсіям.

Порівняння з аналогічними підходами дозволяє зробити висновок про значно кращу ефективність при використанні текстів середнього та великого об'єму. Крім того, ефективність та надійність методу доводить стійкість до спеціалізованих засобів видалення чи знешкодження подібних маркерів.

Практичне значення має методологічна основа, яка може бути впроваджена як у сучасних DLP системах, так і в системах ЕЦП чи для протидії інформаційним впливам.

Подальшого дослідження потребує аналіз впливу кількості модифікованих семантично неподільних складових англомовного тексту на ставлення зловмисника до тексту зворотньої пропаганди. Результати, описані в статті, можуть бути використані при розробці стегаграфічного протоколу для вбудовування ЦВЗ у текст, враховуючи інформаційне протидиверсія.

Список використаних джерел

- [1] Указ Президента України "Про рішення Ради національної безпеки і оборони України від 15 жовтня 2021 року "Про Стратегію інформаційної безпеки України". [Електронний ресурс]. Режим доступу: <https://zakon.rada.gov.ua/laws/show/685/2021#n7>. Дата звернення: Лют. 18, 2022.
- [2] 2020 Cost of Insider Threats Global Report, 2020. [Online]. Available: <https://www.exclusive-networks.com/uk/wp-content/uploads/sites/28/2020/12/UK-VR-Proofpoint-Report-2020-Cost-of-Insider-Threats.pdf>. Accessed on: Febr. 20, 2022.
- [3] S. Chen, J. Wang, X. Feng, F. Jiang, B. Qin, and C.-Y. Lin, "Enhancing neural data-to-text generation models with external background knowledge", in *Proc. 2019 Conf. on Empirical Methods in Natural Language Processing and 9th Int. Joint Conf. on Natural Language Processing*, Hong Kong, China, Nov. 3-7, 2019, pp. 3022-3032.
- [4] S. Welleck, K. Brantley, H. Daume III, and K. Cho, "Non-monotonic sequential text generation", in *Proc. 36th Int. Conf. on Machine Learning*, Long Beach, USA, June 10-15, 2019.
- [5] T. C. Ferreira, C. van der Lee, E. van Miltenburg, and E. Kraemer, "Neural data-to-text generation: A comparison between pipeline and end-to-end architectures", in *Proc. 2019 Conf. on Empirical Methods in Natural Language Processing and 9th Int. Joint Conf. on Natural Language Processing*, Hong Kong, China, Nov. 3-7, 2019, pp. 552-562.
- [6] Y. Yu, D. Qiu, and R. Yan, "A quantum language-inspired tree structural text representation for semantic analysis", *Mathematics*, no. 10 (6), p. 914, 2022.
- [7] S. Y. Feng, A. W. Li, and J. Hoey, "Keep calm and switch on! Preserving sentiment and fluency in semantic text exchange", in *Proc. 2019 Conf. on Empirical Methods in Natural Language Processing and 9th Int. Joint Conf. on Natural Language Processing*, Hong Kong, China, Nov. 3-7, 2019, pp. 2701-2711.
- [8] С. Д. Погорілий, та А. А. Крамов, "Метод розрахунку когерентності українського тексту", *Реєстрація, зберігання і обробка даних*, т. 20, № 4, с. 64-75, 2018.
- [9] Я. В. Тарасенко, "Метод комп'ютеризованого формування англомовного тексту відповідно до психолінгвістичного портрету пропагандиста", *Захист інформації*, т. 22, № 2, с. 66-73, 2020.
- [10] Ya. Tarasenko, "The quantum-semantic psycholinguistic analysis method for the english-language text of propaganda discourse", *Advanced Information Systems*, vol. 3, no. 4, pp. 62-68, 2019.
- [11] Я. В. Тарасенко, "Метод виявлення і збереження неподільної семантичної складової англомовного маніпулятивного тексту пропагандного дискурсу на основі квантового формалізму", *Вісник Черкаського державного технологічного університету. Технічні науки*, № 1, с. 70-78, 2021.
- [12] Я. В. Тарасенко, "Метод додавання семантичного шуму за індивідуальною семантичною лінією пропагандиста", *Безпека інформації*, т. 27, № 2, с. 100-108, 2021.
- [13] C. E. Osgood, and P. H. Tannenbaum, "The principle of congruity in the prediction of attitude change", *Psychological Review*, no. 62 (1), pp. 42-55, 1955.
- [14] Я. В. Тарасенко, "Засоби оцінювання ставлення пропагандиста до тексту зво-

ротного психологічного впливу", *Вісник Вінницького політехнічного інституту*, № 4, с. 78-84, 2021.

- [15] Я. В. Тарасенко, О. Б. Півень, та І. М. Федотова-Півень, "Метод семантичного стиснення текстової інформації для протидії комп'ютерній лінгвістичній стегаграфії", *Наука і техніка Повітряних Сил Збройних Сил України*, № 3 (32), с. 68-78, 2018.

References

- [1] *Decree of the President of Ukraine "On the decision of the National Security and Defense Council of Ukraine on October 15, 2021 "About the Strategy of information security of Ukraine"*. [Online]. Available: <https://zakon.rada.gov.ua/laws/show/47/2017#Text>. Accessed on: Febr. 18, 2022 [in Ukrainian].
- [2] 2020 Cost of Insider Threats Global Report, 2020. [Online]. Available: <https://www.exclusive-networks.com/uk/wp-content/uploads/sites/28/2020/12/UK-VR-Proofpoint-Report-2020-Cost-of-Insider-Threats.pdf>. Accessed on: Febr. 20, 2022.
- [3] S. Chen, J. Wang, X. Feng, F. Jiang, B. Qin, and C.-Y. Lin, "Enhancing neural data-to-text generation models with external background knowledge", in *Proc. 2019 Conf. on Empirical Methods in Natural Language Processing and 9th Int. Joint Conf. on Natural Language Processing*, Hong Kong, China, Nov. 3-7, 2019, pp. 3022-3032.
- [4] S. Welleck, K. Brantley, H. Daume III, and K. Cho, "Non-monotonic sequential text generation", in *Proc. 36th Int. Conf. on Machine Learning*, Long Beach, USA, June 10-15, 2019.
- [5] T. C. Ferreira, C. van der Lee, E. van Miltenburg, and E. Kraemer, "Neural data-to-text generation: A comparison between pipeline and end-to-end architectures", in *Proc. 2019 Conf. on Empirical Methods in Natural Language Processing and 9th Int. Joint Conf. on Natural Language Processing*, Hong Kong, China, Nov. 3-7, 2019, pp. 552-562.
- [6] Y. Yu, D. Qiu, and R. Yan, "A quantum language-inspired tree structural text representation for semantic analysis", *Mathematics*, no. 10 (6), p. 914, 2022.
- [7] S. Y. Feng, A. W. Li, and J. Hoey, "Keep calm and switch on! Preserving sentiment and fluency in semantic text exchange", in *Proc. 2019 Conf. on Empirical Methods in Natural Language Processing and 9th Int. Joint Conf. on Natural Language Processing*, Hong Kong, China, Nov. 3-7, 2019, pp. 2701-2711.
- [8] S. D. Pogorilyy, and A. A. Kramov, "Method of the coherence evaluation of Ukrainian text", *Reiestratsiia, zberihannia i obrobka danykh*, vol. 20, no. 4, pp. 64-75, 2018 [in Ukrainian].
- [9] Ya. V. Tarasenko, "The method of English text's computerized formation in accordance with the propagandist's psycholinguistic portrait", *Zakhyst informatsii*, vol. 22, no. 2, pp. 66-73, 2020 [in Ukrainian].
- [10] Ya. Tarasenko, "The quantum-semantic psycholinguistic analysis method for the english-language text of propaganda discourse", *Advanced Information Systems*, vol. 3, no. 4, pp. 62-68, 2019.
- [11] Ya. V. Tarasenko, "Method of detection and preservation of indivisible semantic component of English manipulative text of propaganda discourse based on quantum formalism", *Visnyk Cherkaskogo derzhavnogo tekhnolohichnogo universytetu. Tekhnichni nauky*, no. 1, pp. 70-78, 2021 [in Ukrainian].
- [12] Ya. V. Tarasenko, "The method of semantic noise addition according to the propagandist's individual semantic line", *Bezpeka informatsii*, vol. 27, no. 2, pp. 100-108, 2021 [in Ukrainian].
- [13] C. E. Osgood, and P. H. Tannenbaum, "The principle of congruity in the prediction of attitude change", *Psychological Review*, no. 62 (1), pp. 42-55, 1955.
- [14] Ya. V. Tarasenko, "Means of assessing the propagandist's attitude to the text of the reverse psychological influence", *Visnyk Vinnytskoho politekhnichnogo instytutu*, no. 4, pp. 78-84, 2021 [in Ukrainian].
- [15] Ya. V. Tarasenko, O. B. Piven and I. M. Fedotova-Piven, "Method of the textual information semantic compression for counteracting computer linguistic steganography", *Nauka i tekhnika Povitrianykh Syl Zbroinykh Syl Ukrainy*, no. 3 (32), pp. 68-78, 2018 [in Ukrainian].

Ya. V. Tarasenko, Ph. D.,
e-mail: yaroslav.tarasenko93@gmail.com
V. H. Babenko, Dr. Tech. Sc., Associate Professor
e-mail: v.babenko@chdtu.edu.ua
Cherkasy State Technological University
Shevchenko blvd, 460, Cherkasy, 18006, Ukraine

METHOD FOR COMPUTERIZED MODIFICATION OF ENGLISH TEXT BASED ON PSYCHOSEMANTIC PROPERTIES

The article investigates the possibility of using psychosemantic properties to protect textual information with limited access. The developed method of computerized modification of English texts allows to increase the effectiveness of preventing the illegal distribution and use of textual information compared to methods that do not take into account the psychosemantic component. This makes it possible to protect modified semantically indivisible elements from their possible distortion or removal, given the threat of potential informational influences. The method allows to protect text documents and counteract cyber sabotage. To achieve the main aim of the article, it is decided to form psychosemantic properties on the basis of categorical apparatus and psycholinguistic profile. This makes it possible to embed the marker into a semantically indivisible component more effectively. In developing this method, the method of calculating the coherence of the text has been also improved and the worlds of perception of the semantic particle have been taken into account. Due to this, it becomes possible to increase the overall level of the text modification efficiency and the naturalness of its perception. The process of computerized modification of English texts with the help of computer tools has been simulated. The results of the comparison with similar approaches allow us to draw preliminary conclusions about the higher efficiency when working with medium and high volume texts. The obtained experimental data also make it possible to determine the resistance of the texts modified by the developed method to the use of specialized means of semantic removal of existing markers. It is revealed that semantic compression had not removed any modified objects. The proposed method can be used in Data Loss Prevention systems and Digital Signature systems. Further development of a steganographic protocol for digital watermark embedding is envisaged with the help of the developed method of English texts modification.

Keywords: *computerized text modification, psychosemantic properties, semantically indivisible component, information with limited access, quantum-semantic analysis, subjective perception of semantics.*

Стаття надійшла 10.03.2022

Прийнято 16.03.2022