

**НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК УКРАИНЫ  
ИНСТИТУТ ПРОБЛЕМ РЕГИСТРАЦИИ ИНФОРМАЦИИ НАН  
УКРАИНЫ**

**ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ  
И БЕЗОПАСНОСТЬ**

**МАТЕРИАЛЫ XIX МЕЖДУНАРОДНОЙ  
НАУЧНО-ПРАКТИЧЕСКОЙ КОНФЕРЕНЦИИ**

**ВЫПУСК 19**

Киев – 2019

*Рекомендовано к печати Ученым советом  
Института проблем регистрации информации НАН Украины  
(протокол № 3 от 24 декабря 2019 г.)*

**Информационные технологии и безопасность. Материалы XIX  
Международной научно-практической конференции ИТБ-2019.** – К.:  
ООО "Инжиниринг", 2019. – 236 с. ISBN 978-966-2344-72-1

В сборник вошли материалы докладов, представленных на XVI  
Международной научно-практической конференции «Информационные  
технологии и безопасность» (ИТБ-2019, 28 ноября 2019 года, г. Киев,  
Украина).

В сборнике представлены статьи, посвященные вопросам  
безопасности живучести критических инфраструктур, моделирования и  
противодействия информационным операциям, информационных  
технологий в управлении, методов и способов информационной  
поддержки принятия решений, компьютерного моделирования систем  
организационного управления, информационно-аналитических  
исследований на основе открытых источников информации, сценарного  
анализа при обеспечения информационной поддержки принятия  
решений, актуальным проблемам обеспечения информационной и  
кибернетической безопасности.

Для специалистов в области информационных технологий,  
информационной безопасности, информационного права а также для  
аспирантов и студентов старших курсов высшей школы  
соответствующих специальностей.

***Редакционная коллегия:***

*А.Г. Додонов, д.т.н., профессор; В.В. Голенков, д.т.н., профессор;  
Минглей Фу, PhD; Д.В. Ландэ, д.т.н., профессор; В.В. Мохор, член-корр  
НАН Украины, д.т.н., профессор; В.В. Хаджинов, д.т.н., профессор;  
В.В. Цыганок, д.т.н., с.н.с.; Е.С. Горбачик, к.т.н., с.н.с.; М.Г. Кузнецова,  
к.т.н., с.н.с., О.В. Андрейчук, к.т.н.*

ISBN 978-966-2344-72-1

© Институт проблем регистрации  
информации НАН Украины, 2019

© Коллектив авторов, 2019

## Conclusions

In this article the problems of gathering flow information scenario optimization on the branched network are considered as an example of the construction and further optimization problem of scenarios built on the branched network for the budget monitoring. The structure approach in ontology analysis, as the most appropriate for the considered task, using the graph representation of the structure is suggested. The estimation of the effect of the previously obtained partial information about gathering information in the network scenario construction, described by the graph, for the further clarification of information. The formal description of multilayer hierarchical system structure is provided. The example of ontology elements interaction structure for the problem is presented.

The complex approach based on the shortest route search on the graph and ontology model graph representation is suggested. This makes possible to use algorithms, based on graph nodes in hierarchical levels (layers) traversing.

The approach for modified search in width algorithm building is presented, which significantly decreases routes search for the gathering flow information scenario construction time in the branched network. Described in the article approach for the gathering flow information scenario optimization on the branched network was tested for the pilot system of regional budgets financial analysis project development. Algorithm suggested here makes it possible to develop the software complex, which enables sufficiently full and complete solution to the problems of scenario optimization for scenarios of search and collection of streaming information on an extensive network. One of the perspective directions in the algorithm application is to use its possibilities for information-analytic systems building. This will significantly reduce the time and improve the quality of searching the necessary streaming information on a extensive network.

## Literature

1. Alex Guazzelli, Michael Zeller, Wen-Ching Lin and Graham Williams PMML: An Open Standard for Sharing Models: available at: [https://journal.r-project.org/archive/2009-1/RJournal\\_2009-1\\_Guazzelli + et+al.pdf](https://journal.r-project.org/archive/2009-1/RJournal_2009-1_Guazzelli_et+al.pdf)
2. Chernov, V.A. The Economic Theory analysis: Textbook // V.A. Chernov .- Moscow: Prospect, 2017 .- 384 p. - ISBN 978-5-392-24867-4
3. Christopher J. Manning, Prabhakar Rahavan, Heinrich Schütz. Introduction Consumer Information Search (trans. With Eng.) - M .: ООО 'Y.D. Williams' 2011 – p.504.
4. O. Koval, V. Kuzminykh, S. Otrokh and V. Kravchenko, "Optimization of Scenarios for Collecting Information Streaming Wide-Area Network," 2019 3rd International Conference on Advanced Information and Communications Technologies (AICT), Lviv, Ukraine, 2019, pp. 213-215. doi: 10.1109/AIACT.2019.8847832.
5. Miller G. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. The Psychological Review, 1956. 63: pp.81-97.
6. Thorup, Mikkel. "Undirected Single-Source Shortest Paths with Positive Integer Weights in Linear Time." Journal of the ACM 46, no. 3 (May 1, 1999): pp.362–394
7. Moore, Edward F. "The Shortest Path Through a Maze". International Symposium on the Theory of Switching, pp.285–292, 1959.
8. Lee, C Y. "An Algorithm for Path Connections and Its Applications." IEEE Transactions on Electronic Computers 10, no. 3 (September 1961): pp. 346–365.

## ВИЗНАЧЕННЯ НАПРЯМКІВ ЗВ'ЯЗКІВ У МЕРЕЖІ ТЕРМІНІВ

Д.В. Ланде<sup>1,2,3[0000-0003-3945-1178]</sup>, О.О. Дмитренко<sup>1[0000-0001-8501-5313]</sup>, О.Г. Радзієвська<sup>3[0000-0003-3813-3987]</sup>,

А.В. Бойченко<sup>1</sup>

<sup>1</sup> Інститут проблем реєстрації інформації НАН України, Київ, Україна

<sup>2</sup> Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", Київ, Україна

<sup>3</sup> Науково-дослідний інститут інформатики і права НАПрН України, Київ, Україна

*dwlande@gmail.com, dmytrenko.o@gmail.com, radeoksa@gmail.com*

**Анотація.** Розвиток комп'ютерних технологій та, зокрема, мережі Інтернет, яка є джерелом інформаційних ресурсів, які, в свою чергу, є динамічним джерелом текстів, відкриває нові можливості для розробки й застосування покращених методів їх дослідження. Існують різні методи, методики та техніки комп'ютеризованої обробки та аналізу тексту. Проте зважаючи на неспинний розвиток, сучасне програмне забезпечення все більше потребує готових рішень для вдосконалення своїх систем. Адже досі

відкритою та до кінця не вирішеною є задача вибору окремих термінів з корпусу текстових документів й автоматизація такого відбору. У зв'язку зі складністю природної мови, також не менш складною й відкритою проблемою є встановлення синтаксичних зв'язків між термінами у тексті, та визначення напрямків таких зв'язків. В даній роботі розглядаються та аналізуються підходи до побудови мереж термінів, як онтологічної моделі предметної області. Зокрема, представлені нові підходи до визначення напрямків зв'язків між вузлами ненаправленої мережі, побудованої зі слів та словосполучень (окремих уніграм, біграм та триграм) тематичного текстового масиву. Також для побудови направленої мережі термінів розглядається й застосовується один із методів створення термінологічних онтологій – алгоритм формування мереж природних ієрархій термінів на основі аналізу текстового корпусу. Для створення програмної реалізації запропонованих та розглянутих підходів і методів використовується мова програмування Python, а також окремі функції спеціалізованої надбудови – модуля NLTK (Natural Language Toolkit open source library). Використовуючи засоби програмного забезпечення для моделювання та візуалізації графів – Gephi, побудовані направлені мережі термінів були візуалізовані з метою кращого наочного сприйняття.

**Ключові слова:** Предметна Область, Онтологія, Мережа Термінів, Граф Горизонтальної Видимості, Мережа Природної Ієрархії Термінів, Синтаксичні та Семантичні Зв'язки, Ненаправлена Мережа, Направлена Мережа.

## 1. Постановка проблеми

Дуже важливим етапом у комплексних дослідженнях певної предметної області (Subject Domain) є детальне формалізоване представлення її знань, придатне для автоматизованої обробки – побудова онтологічної моделі предметної області. В якості моделі предметної області може розглядатися мережа термінів (Network of Terms), вузли якої відповідають окремим словам та словосполученням у тексті, а ребра – зв'язкам між ними. Процес побудови великих тематичних онтологій зазвичай є складним та ресурсозатратним та становить нерозв'язану досі науково-практичну проблему [1]. Окремий крок такої формалізації – це визначення базових об'єктів. У випадку побудови мереж термінів – це створення словникових номенклатур, тезаурусів та предметних словників термінів, визначених на основі корпусу текстових документів. Ефективний вибір окремих термінів й, тим більше, автоматизація такого відбору з тематичного текстового масиву – актуальна й до кінця не вирішена задача [2, 3].

У зв'язку зі складністю природної мови, не менш складною й відкритою проблемою концептуалізації є встановлення семантико-синтаксичних зв'язків між вузлами мережі, що відповідають термінам, та визначення напрямків таких зв'язків.

Метою даної роботи є представити нові підходи до визначення напрямків зв'язків між вузлами ненаправленої мережі, побудованої зі слів та словосполучень тематичного текстового масиву.

## 2. Метод побудови ненаправлених мереж термінів

Для побудови мереж термінів існує декілька підходів і способів інтерпретації вузлів та зв'язків [4, 5], що призводить до різних видів представлення таких мереж [6].

Для побудови термінологічних онтологій предметних областей в даній роботі використовується алгоритм компактифікованого графа горизонтальної видимості для ключових термінів: окремих уніграм, біграм та триграм.

### 2.1. Компактифікований граф горизонтальної видимості

Граф горизонтальної видимості [7, 8, 9] є модифікацією стандартного алгоритму графа видимості [10].

Для того, щоб побудувати ненаправлену мережу термінів з використанням алгоритму горизонтальної видимості у роботі [11] пропонується здійснити наступні кроки. Перший етап полягає у тому, що на горизонтальній осі відмічається ряд вузлів, кожен з яких відповідає термінам у тому порядку, в якому вони з'являються в тексті; а по вертикальній осі відкладаються вагові значення – числові оцінки  $x_i$ . На другому етапі будується граф горизонтальної видимості. Вважається, що два вузли  $t_i$  та  $t_j$ , які відповідають елементам часового ряду  $x_i$  і  $x_j$ , знаходяться у горизонтальній видимості тоді й тільки тоді, коли  $x_k < \min(x_i; x_j)$  для всіх  $t_k$  таких, що  $t_i < t_k < t_j$ .

Третій етап полягає в тому, що отримана на попередніх етапах мережа компактифікується: вузли, що відповідають однаковим термінам, об'єднуються в один вузол. Отримана ненаправлена мережа термінів буде називатись компактифікованим графом горизонтальної видимості (рис. 1).

Таким чином, алгоритм компакфікованого графа горизонтальної видимості дозволяє будувати ненаправлені мережеві структури на основі текстів у випадку, коли окремим словам або словосполученням поставлені у відповідність числові вагові значення.

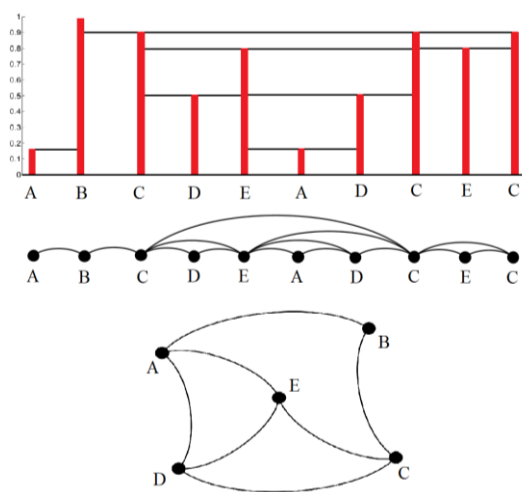


Рис. 12. Етапи побудови компакфікованого графу горизонтальної видимості [11].

## 2.2. Обробка текстового корпусу

Усна та писемна форми мови складаються зі слів, які часто мають спільне походження.

Мова, що містить різні форми слова, похідні від іншого слова, які використовуються для вираження різного змісту, називається переплетеною мовою (Inflected Language). Зрозуміло, що такі словоформи мають спільну основу.

В даній секції представлені основні етапи комп'ютеризованого процесу обробки текстових документів, які були використані в роботі, такі як: токенізація, розбір на частини мови, лематизація, вилучення стоп-слів, стемінг та зважування термінів.

### *Токенізація та лематизація*

Для проведення попереднього лексичного аналізу здійснюється процес розбиття тексту на елементарні одиниці (токени, лексеми) – токенізація та подальша їх лематизація. Лематизація – процес приведення словоформи до леми – її нормальної (словникової) форми.

У цій роботі була використана бібліотека Python – NLTK, а саме “WordNet Lemmatizer”

Токенізація та лематизація є зазвичай початковими етапами обробки текстів, адже дають змогу працювати зі словом як з окремою сутністю, при цьому знаючи його контекст [12].

### *Розбір на частини мови*

Розбір на частини мови є одним з перших етапів комп'ютерного аналізу тексту.

Для того, щоб кожному слову та його словоформі була надана правильна початкова форма у результаті лематизації, необхідно спочатку надати їм контекст, здійснивши розбір на частини мови [13].

У корпусній лінгвістиці, розбір на частини мови (англ. part-of-speech tagging, POS tagging) – це процес віднесення слова в тексті (корпусі) до певної частини мови, заснований як на його визначенні, так і на його контексті – тобто, на його зв'язку з суміжними і спорідненими словами у фразі, реченні, або абзаці.

Алгоритми розбору на частини мови поділяються на дві відмінні групи: на основі правил і на стохастичні. Метод розбору Е. Брілла [14], який використовує алгоритми на основі правил, є першим й найбільш використовуваним методом розбору англійських текстів.

### *Вилучення стоп-слів*

Також після етапу обробки текстових документів та виокремлення ключових термінів в даному дослідженні пропонується вилучити стоп-слова, які не мають ніякого смислового навантаження, тобто є інформаційно-неважливими, а також біграми, які містять принаймні одне стоп-слово, та триграми, які починаються, або закінчуються стоп-словом. Стоп-словник, який використовувався в межах даної роботи, був сформований на основі різних стоп-словників, які доступні за посиланнями:

<https://code.google.com/archive/p/stop-words/downloads/>;

<http://www.textfixer.com/tutorials/common-english-words.php>.

Для створення програмної реалізації додатково було використано “SnowballStemmer” – функцію модуля NLTK (Natural Language Toolkit) – спеціалізованої надбудови мови програмування Python.

Також сформований стоп-словник доповнювався іншими стоп-словами, які були виявлені експертами в межах досліджуваної області.

### **Стемінг**

Після етапів, описаних вище, щоб об'єднати слова, які мають спільний корінь, з метою нормалізації тексту, пропонується здійснити процес стематизації – скорочення слова до основи шляхом відкидання афіксів (закінченєвих змінних морфем-афіксів, постфіксів, суфіксів та префіксів), що формують похідні форми слова [15]. Результати стемінгу іноді дуже схожі на визначення кореня слова, але його алгоритми базуються на інших принципах [16]. Тому слово після стематизації (обробки алгоритмом стемінгу) може відрізнитися від морфологічного кореня слова.

Варто відзначити, що лематизація та стемінг мають на меті звести словозмінну форму слова, або інколи його похідну споріднену форму, до загальної основної форми.

Проте ці два процеси відрізняються тим, що стемінг зазвичай згортає похідні форми споріднених слів, скорочуючи їх до основи; тоді як лематизація зазвичай лише руйнує різні похідні словоформи, приводячи до нормальної (словникової) форми – леми.

Якщо розглядати англійське слово "saw", стемінг може повернути лише "s", хоча лематизація намагатиметься повернути або "see", або "saw" залежно від того, чи використовувалась лексема (токен) як дієслово чи як іменник.

Тому, щоб уникнути ситуації, описаної вище, в даній роботі процес лематизації проводиться перед стематизацією.

Існує декілька типів алгоритмів стемінгу, які розрізняються відносно продуктивності, точності та відносно того, як долаються проблеми стемінгу [17].

В даній роботі, для створення програмної реалізації, було використано стример “PorterStemmer” (окрема функція спеціалізованої надбудови – модуля NLTK (Natural Language Toolkit)), що розроблений на мові програмування Python та реалізує алгоритм стемінгу Мартіна Портера [18, 19]. Алгоритм Портера є де-факто стандартним алгоритмом стемінгу для англійської мови.

Описані вище кроки дають змогу нормалізувати текст корпусу.

## **2.3. Зважування та виокремлення ключових термінів**

Після проведення попередніх етапів обробки текстового корпусу здійснюється процес зважування та виокремлення ключових термінів. В якості вагових значень термінів, для формування часового ряду в якості функції, яка ставить у відповідність терміну число, в даній роботі використовується модифікація класичного статистичного вагового показника важливості терміна TF-IDF (з англ. Term Frequency – частота терміна, Inverse Document Frequency – обернена частота документа) [20, 21], а саме – GTF (Global Term Frequency – глобальна частота терміна) [22].

Цей підхід дозволяє інформаційно-важливим в глобальному контексті елементам тексту мати високий статистичний показник важливості.

## **3. Визначення напрямків зв'язків**

Нехай  $G$  – ненаправлена мережа термінів побудована за принципом, що описаний вище:  $G := (V, T)$  де  $V$  — множина вузлів,  $T$  — множина неупорядкованих пар вузлів з  $V$ , які відповідають причинно-наслідковим зв'язкам між вузлами.

Вважається, що  $\forall_{i,j} : (t_i, t_j) \in T$  причинно-наслідковий зв'язок існує у напрямку від вузла  $t_i$  до  $t_j$  якщо:

1) числове значення, що відповідає: а) степеню [23, 24] б) показнику, що обчислюється за алгоритмом HITS [25] в) показнику, що обчислюється за алгоритмом PageRank [26]) вузла  $t_i$  більше за числове значення вузла  $t_j$  ;

2) у реченні термін, якому відповідає вузол  $t_i$  зустрічається раніше терміна, якому відповідає вузол  $t_j$  ;

3) термін, якому відповідає  $t_i$  коротший за термін, якому відповідає  $t_j$  .

Для побудови направленої мережі зі слів та словосполучень (уніграм, біграм та триграм) за третім правилом використовується один із методів створення термінологічних онтологій – алгоритм формування мереж природніх ієрархій термінів. Як зазначено у роботі [27] алгоритм створення мереж природніх ієрархій термінів передбачає побудову компактифікованого графу горизонтальної

видимості, та встановлення направлених зв'язків між ключовими термінами, де напрямок зв'язку визначається за принципом входження слова у двочленне або тричленне словосполучення, та входження двочленного у тричленне.

#### 4. Результати дослідження запропонованого підходу

Запропонований підхід для визначення напрямків зв'язків у ненаправлених мережах термінів був апробований на прикладі англomовного тексту, а саме – відомої казки “The story of Little Red Riding Hood”.

Відповідно до вищеописаного методу було здійснено обробку обраного текстового документу й виокремлено ключові терміни: уніграми, біграми та триграми (Таблиця 1).

**Таблиця 2.** Топ-26 ключових уніграм для тексту “The story of Little Red Riding Hood” і їх степінь, HITS та PageRank.

Уніграми	GTF	Степінь	HITS	PageRank
grandmoth	0.065	49	0.444	0.0545
red	0.059	32	0.3099	0.036
hood	0.053	22	0.256	0.0252
ride	0.053	2	0.051	0.0029
wolf	0.031	30	0.301	0.0327
wood	0.025	17	0.204	0.0169
bed	0.019	18	0.191	0.0186
open	0.016	13	0.19	0.0148
time	0.016	14	0.199	0.0162
beauti	0.016	15	0.155	0.0154
big	0.012	9	0.088	0.0119
flower	0.012	8	0.126	0.0093
door	0.012	10	0.136	0.0125
cap	0.012	12	0.162	0.0141
cake	0.012	9	0.101	0.0116
wine	0.012	7	0.099	0.0094
cut	0.009	10	0.095	0.0127
mother	0.009	7	0.118	0.0092
strang	0.009	13	0.116	0.0167
jump	0.009	9	0.101	0.0116
weak	0.009	6	0.083	0.0083
ate	0.009	7	0.134	0.0081
live	0.009	6	0.097	0.0073
pull	0.009	7	0.12	0.0083
sick	0.009	4	0.05	0.0059
hunter	0.009	11	0.113	0.0148

**Таблиця 3.** Топ-22 ключові біграми для тексту “The story of Little Red Riding Hood” і їх степінь, HITS та PageRank.

Біграми	GTF	Степінь	HITS	PageRank
ride_hood	0.053	26	0.465	0.0277
red_ride	0.053	28	0.494	0.0303
grandmoth_big	0.009	11	0.177	0.0095
hood_wolf	0.006	5	0.179	0.0054
tasti_bite	0.006	8	0.192	0.0085
bed_pull	0.053	3	0.021	0.0038
hood_grandmoth	0.053	4	0.107	0.0047
leav_path	0.009	7	0.162	0.0084

snore_loudli	0.006	8	0.055	0.0068
grandmoth_live	0.006	6	0.164	0.0071
wolf_bodi	0.006	7	0.160	0.0080
pull_curtain	0.006	5	0.106	0.0056
grandmoth_bed	0.006	5	0.04	0.0060
straight_grandmoth	0.006	7	0.133	0.0076
sick_weak	0.006	5	0.068	0.0057
beauti_wood	0.006	6	0.133	0.0070
cake_wine	0.006	8	0.172	0.0084
beauti_flower	0.006	6	0.105	0.0074
wood_wolf	0.006	7	0.193	0.0076
press_latch	0.006	8	0.047	0.0064
grandmoth_sick	0.006	5	0.096	0.0058
door_open	0.006	8	0.109	0.0085

**Таблиця 4.** Топ-26 ключових триграм для тексту “The story of Little Red Riding Hood” і їх степінь, HITS та PageRank.

Триграми	GTF	Степінь	HITS	PageRank
red_ride_hood	0.1429	36	0.67	0.1042
grandmoth_what_big	0.0252	6	0.145	0.0114
press_the_latch	0.0168	6	0.059	0.0111
leav_the_path	0.0168	4	0.153	0.0126
sick_and_weak	0.0168	6	0.182	0.0179
bed_and_pull	0.0168	7	0.162	0.0188
cake_and_wine	0.0168	5	0.160	0.0133
hear_how_beauti	0.0084	2	0.111	0.0076
look_so_strang	0.0084	2	0.03	0.0071
hood_and_ate	0.0084	2	0.111	0.0079
listen_littl_red	0.0084	2	0.129	0.007
bite_he_climb	0.0084	2	0.001	0.0101
obey_her_mother	0.0084	2	0.021	0.0084
lay_the_wolf	0.0084	2	0	0.0105
mind_your_manner	0.0084	2	0.054	0.0068
open_hi_belli	0.0084	2	0.003	0.0096
cake_and_drunk	0.0084	2	0.018	0.0090
snore_veri_loudli	0.0084	2	0	0.0105
strang_oh_grandmoth	0.0084	2	0.028	0.0059
bird_are_sing	0.0084	2	0.021	0.0084
bed_fell_asleep	0.0084	2	0	0.0103
ride_hood_enter	0.0084	2	0.114	0.0074
larg_heavi_stone	0.0084	2	0.004	0.0093
woman_wa_snore	0.0084	2	0.111	0.0076
loudli_a_huntsman	0.0084	2	0	0.0105
red_ride_hood	0.0084	2	0	0.1042

Побудувавши направлену мережу за першим правилом було отримано наступні результати для різних числових показників вузлів мережі ( а) для степеня вузла – рис. 2; б) для HITS – рис. 3; в) для PageRank – рис.4). Візуалізація побудованих мереж термінів здійснювалась за допомогою засобів програмного забезпечення для моделювання та візуалізації графів – Gephi (<https://gephi.org>).



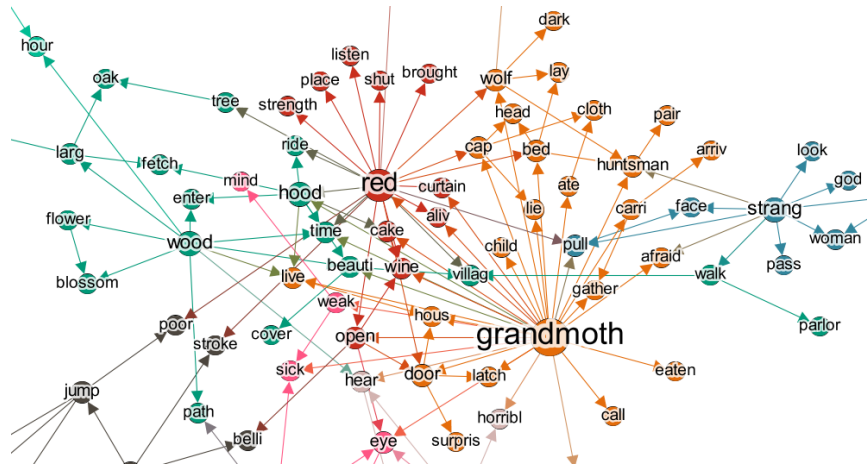


Рис. 13. Фрагмент направленої мережі побудованої за першим правилом для а) степеня вузла.

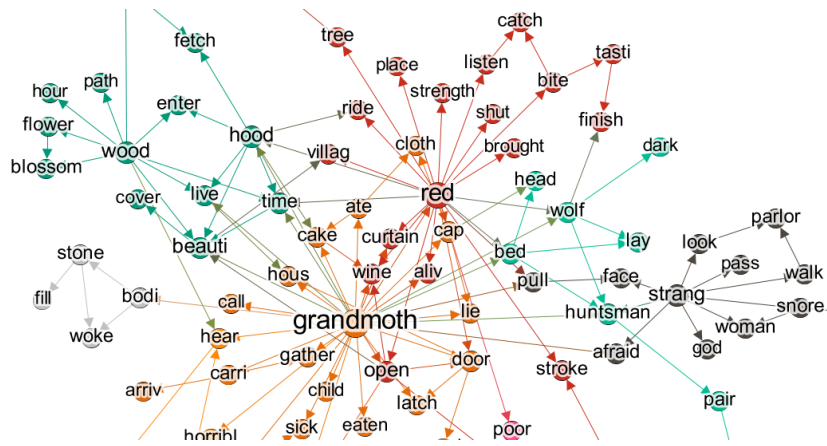


Рис. 14. Фрагмент направленої мережі побудованої за першим правилом для б) HITS.

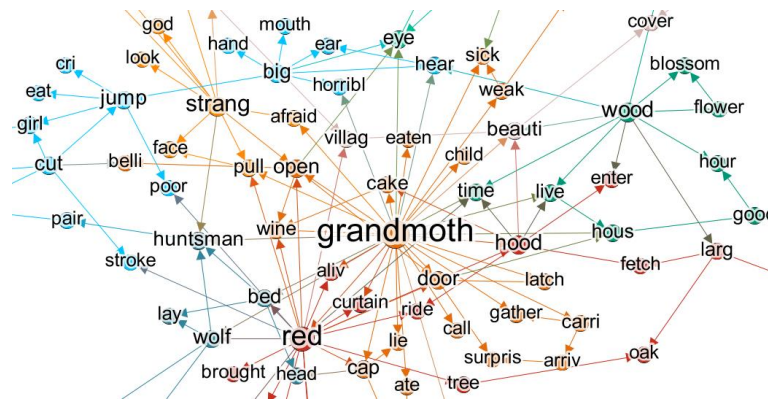


Рис. 15. Фрагмент направленої мережі побудованої за першим правилом для в) PageRank.

На рис. 5 представлена направлена мережа термінів, що побудована за другим правилом.



of Little Red Riding Hood” було побудовано ненаправлену мережу термінів, метод побудови якої був представлений у даній роботі. Застосовуючи запропоновані правила визначення напрямків зв'язків, із ненаправленої мережі термінів було отримано направлені мережі. Змістовність зв'язків для мережі побудованої за другим запропонованим правилом є найвищою серед інших двох правил на думку експертів. Зважаючи на природність зв'язків, які встановлюються у мережі природних ієрархій термінів, можна говорити про їх синтаксичну адекватність.

Направлені мережі зі слів та словосполучень, побудовані за допомогою запропонованого підходу, можна використовувати як основу для автоматизованої побудови онтологічних моделей предметних областей. Також результати роботи можуть бути використані під час створення персональних пошукових інтерфейсів для користувачів інформаційно-пошукових систем, а також у системах навігації у базах даних. Це повинно допомогти користувачам таких систем спростити процес пошуку релевантної інформації.

Оскільки задача підвищення точності визначення напрямків зв'язків у ненаправлених мережах зі слів та словосполучень залишається актуальною, планується продовжити роботу в цьому напрямку, розвиваючи нові та модифікуючи існуючі підходи.

## Літературні джерела

1. Ландэ, Д.В., Снарский, А.А.: Подход к созданию терминологических онтологий. Онтология проектирования 2(12). 83-91 (2014).
2. Лукашевич, Н.В., Добров, Б., Чуйко, Д.С.: Отбор словосочетаний для словаря системы автоматической обработки текстов. Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог–2008», С. 339–344. М. (2008).
3. Филиппович, Ю.Н., Прохоров, А.В.: Семантика информационных технологий: Опыты словарно-тезаурусного описания. М., МГУП (2002).
4. Ferrer-i-Cancho, R., & Solé, R.: The Small World of Human Language. in Proc. of the Royal Society of London, pp. 2261-2265. London (2001).
5. doi: 10.1098/rspb.2001.1800.
6. Caldeira, S. M. G., Petit Lobao, T. C., Andrade, R. F. S., Neme, A., & Miranda, J. G. V.: The network of concepts in written texts. The European Physical Journal B-Condensed Matter and Complex Systems 49(4), 523-529 (2005).
7. Ferrer-i-Cancho, R. F., Solé, R. V., & Köhler, R.: Patterns in syntactic dependency networks. Physical Review E 69(5), (2004).
8. doi: 10.1103/PhysRevE.69.051915
9. Luque, B., Lacasa, L., Ballesteros, F., & Luque, J.: Horizontal visibility graphs: Exact results for random time series. Physical Review E, 80(4), (2009).
10. doi: 10.1103/PhysRevE.80.046103.
11. Gutin, G., Mansour, T., & Severini, S.: A characterization of horizontal visibility graphs and combinatorics on words. Physica A: Statistical Mechanics and its Applications, 390(12), 2421-2428 (2011).
12. doi: 10.1016/j.physa.2011.02.031.
13. Bezudnov, I. V., & Snarskii, A. A.: From the time series to the complex networks: The parametric natural visibility graph. Physica A: Statistical Mechanics and its Applications, 414, 53-60 (2014).
14. doi: 10.1016/j.physa.2014.07.002.
15. Lacasa, L., Luque, B., Ballesteros, F., Luque, J., & Nuno, J. C.: From time series to complex networks: The visibility graph. Proceedings of the National Academy of Sciences, 105(13), 4972-4975 (2008).
16. doi: 10.1073/pnas.0709247105
17. Lande, D. V., Snarskii, A. A., Yagunova, E. V., & Pronoza, E. V.: The use of horizontal visibility graphs to identify the words that define the informational structure of a text. In: 2013 12th Mexican International Conference on Artificial Intelligence, pp. 209-215 (2013).
18. Manning, C. D., Raghavan, P., & Schütze, H.: An Introduction to Information Retrieval. Cambridge University Press, 22–36 (2009).
19. Schmid, H.: Probabilistic Part-of-Speech Tagging Using Decision Trees. In: Proceedings of International Conference on New Methods in Language Processing, pp. 1–9. Manchester, UK (1994).
20. Brill, E.: A simple rule-based part of speech tagger. In: Proceedings of the third conference on Applied natural language processing (ANLC '92). Association for Computational Linguistics, pp. 152-155. Stroudsburg, PA. USA (1992).
21. doi:10.3115/974499.974526

22. Jongejan, B., & Dalianis, H.: Automatic training of lemmatization rules that handle morphological changes in pre-, in-and suffixes alike. In Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, pp. 145-153. Association for Computational Linguistics, Singapore (2009)
23. Lovins, J. B.: Development of a stemming algorithm. *Mech. Translat. & Comp. Linguistics* 11(1-2), 22-31 (1968).
24. Baeza-Yates, R., & Ribeiro-Neto, B.: *Modern information retrieval*. New York: ACM Press, Harlow, England: Addison-Wesley. (2011).
25. Porter, M. F.: An algorithm for suffix stripping. *Program* 14(3), 130-137 (1980).
26. doi: 10.1108/eb046814
27. Willett, P.: The Porter stemming algorithm: then and now. *Program* 40(3), 219-223 (2006).
28. doi: 10.1108/00330330610681295.
29. Salton, G., & Buckley, C.: Term-weighting approaches in automatic text retrieval. *Information processing & management* 24(5), 513-523 (1988).
30. doi:10.1016/0306-4573(88)90021-0
31. Rajaraman, A., & Ullman, J. D. *Mining of massive datasets*. Cambridge University Press (2011).
32. Lande, D.V., Dmytrenko, O.O., & Snarskii A.A.: Transformation texts into complex network with applying visibility graphs algorithms. In: *CEUR Workshop Proceedings (ceur-ws.org)*. Vol-2318 urn:nbn:de:0074-2318-4. Selected Papers of the XVIII International Scientific and Practical Conference on Information Technologies and Security (ITS 2018). vol. 2318. pp. 95-106. (2018).
33. Bondy, J. A., & Murty, U. S. R.: *Graph theory with applications*. vol. 290. Macmillan, London (1976).
34. Godsil, C., & Royle, G.: *Algebraic Graph Theory*. Graduate Texts in Mathematics 207. Springer, New York (2001).
35. doi: 10.1007/978-1-4613-0163-9
36. Kleinberg, J. M.: Authoritative sources in a hyperlinked environment. 26. In *Processing of ACM-SIAM Symposium on Discrete Algorithms*, 46(5), pp. 604–632 (1998).
37. Brin, S., & Page, L.: The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7), 107-117 (1998).
38. doi:10.1016/S0169-7552(98)00110-X
39. Lande, D.V.: Building of networks of natural hierarchies of terms based on analysis of texts corpora. arXiv preprint arXiv:1405.6068 (2014).

## PROBABILISTIC CRITERION OF INFORMATION SECURITY MANAGEMENT SYSTEM DEVELOPMENT

Volodymyr Mokhor<sup>1</sup>[0000-0001-5419-9332]  
 Oleksandr Bakalynskyi<sup>2</sup>[0000-0001-9712-2036],  
 Vasyl Tsurkan<sup>3</sup>[0000-0003-1352-042X]

<sup>1</sup> *Pukhov institute for modeling in energy engineering  
of National academy of sciences of Ukraine, Kyiv-164, Ukraine*

<sup>2</sup> *Department of formation and implementation of state policy on cyber protection  
of Administration of State serves of special communication and information protection  
of Ukraine, Kyiv-110, Ukraine*

<sup>3</sup> *Institute of special communication and information protection National technical University of Ukraine  
“Igor Sikorsky Kyiv Polytechnic Institute”,  
Kyiv-056, Ukraine*

*v.mokhor@gmail.com, baov@meta.ua, v.v.tsurkan@gmail.com*

**Abstract.** The risk assessment presentation features of information security by the risk maps are considered. The attention is paid to special aspects of such presentation. Established limit for discreteness and unevenness of changing the step the value of a risk. Using of continuous risk maps is proposed to overcome the limit. This is due to income the information security events with a continuous flow to consider the analogy information security management system with the queuing system. Therefore the justification of continuous risk map is led for probability of occurrence events with risk acceptance. For this the concept and methods of geometric probability are used. Through this received “unit square” as a reflection of probabilistic geometry which is corresponding of the value normalized quantity of information security risk. The admissibility limit is represented by a hyperbola. With this in mind, use the continuous risk maps is justified.

## СОДЕРЖАНИЕ

<i>Dodonov O.G., Gorbachyk O.S., Kuznietsova M.G.</i> <b>SURVIVABILITY OF ORGANIZATIONAL MANAGEMENT SYSTEMS AND THE MAINTENANCE OF CRITICAL INFRASTRUCTURE SECURITY.....</b>	3
<i>Antonishyn M., Misnik O.</i> <b>ANALYSIS OF TESTING APPROACHES TO ANDROID MOBILE APPLICATION VULNERABILITIES.....</b>	9
<i>Balagura I., Kadenko S., Andriichuk O., and Gorbov I.</i> <b>DEFINING POTENTIAL ACADEMIC EXPERT GROUPS BASED ON JOINT AUTHORSHIP NETWORKS USING DECISION SUPPORT TOOLS.....</b>	17
<i>Beliak Ie.V., Kryuchyn A.A.</i> <b>DEVELOPMENT OF THE MULTISPECTRAL VOLUME RECORDING METHODS.....</b>	18
<i>Berkman L., Otrokh S., Kuzminykh V., Hryshchenko O.</i> <b>METHOD OF FORMATION SHIFT INDEXES VECTOR BY MINIMIZATION OF POLYNOMIALS.....</b>	25
<i>Chertov O. Malchykov V.</i> <b>RATIONAL WAVELET TRANSFORM WITH REDUCIBLE RATIONAL DILATION FACTOR.....</b>	32
<i>Dodonov A., Nikiforov A., Putyatin V., Dodonov V.</i> <b>MODELING COMPLEXES OF ORGANIZATIONAL MANAGEMENT AUTOMATED SYSTEMS - A MEANS TO OVERCOME THE MANAGEMENT CRISIS.....</b>	37
<i>Gladun A., Rogushina J.</i> <b>ЗАСТОСУВАННЯ ОНТОЛОГІЧНОГО АНАЛІЗУ ДЛЯ ОБРОБКИ ВЕЛИКИХ ДАНИХ У ДОМЕНІ КІБЕРБЕЗПЕКИ.....</b>	49
<i>Горбатенко А., Антощук С.</i> <b>ІНФОРМАЦІЙНА ПІДТРИМКА ЛЮДЕЙ З ПРОБЛЕМАМИ ЗОРУ НА ОСНОВІ МІКРОХВИЛЬОВОГО РАДАРУ AWR 1843.....</b>	58
<i>Havrylovych M., Kuznietsova N.</i> <b>SURVIVAL ANALYSIS METHODS FOR CHURN PREVENTION IN TELECOMMUNICATIONS INDUSTRY.....</b>	66
<i>Kadenko S., Tsyganok V., Karabchuk A.</i> <b>COMPARING EFFICIENCY OF EXPERT DATA AGGREGATION METHODS.....</b>	76
<i>Korniyenko B.Y., Galata L. P., Ladieva L.R.</i> <b>MATHEMATICAL MODEL OF THREATS RESISTANCE IN THE CRITICAL INFORMATION RESOURCES PROTECTION SYSTEM.....</b>	86
<i>Костенко Н.Г., Броховецький І.В., Баришполь Д.В.</i> <b>ЗАХИСТ ТА ПІДВИЩЕННЯ ЖИВУЧОСТІ КРИТИЧНИХ СТРУКТУР: ЗАРУБІЖНИЙ ДОСВІД ТА МОЖЛИВОСТІ ДЛЯ УКРАЇНИ.....</b>	92
<i>Koval O.V., Kuzminykh V.O., Voronko M.P.</i> <b>STANDARD ANALYTIC ACTIVITY SCENARIOS OPTIMIZATION BASED ON SUBJECT AREA ANALYSIS.....</b>	98
<i>Ланде Д.В., Дмитренко О.О., Радзівєвська О.Г.</i> <b>ВИЗНАЧЕННЯ НАПРЯМКІВ ЗВ'ЯЗКІВ У МЕРЕЖІ ТЕРМІНІВ.....</b>	103
<i>Mokhor V., Bakalynskiy O., Tsurkan V.</i> <b>PROBABILISTIC CRITERION OF INFORMATION SECURITY MANAGEMENT SYSTEM DEVELOPMENT.....</b>	112
<i>Rogushina J.V.</i> <b>USE OF SEMANTIC SIMILARITY ESTIMATES FOR UNSTRUCTURED DATA ANALYSIS...</b>	118
<i>Rogushina J., Gladun A., Pryima S., Strokan O.</i> <b>ONTOLOGY-BASED APPROACH TO VALIDATION OF LEARNING OUTCOMES FOR INFORMATION SECURITY DOMAIN.....</b>	126
<i>Шнурко-Табаківа Е.В., Ланде Д.В.</i> <b>МЕТОДИ І ЗАСОБИ ІНФОРМАЦІЙНО-АНАЛІТИЧНОЇ ПІДТРИМКИ ПРОТИДІЇ ГІБРИДНИМ ЗАГРОЗАМ ДЕРЖАВИ.....</b>	136
<i>Снарський А.О., Ланде Д.В., Дмитренко О.О.</i> <b>ПОКАЗНИК РЕЛАКСАЦІЇ В СКЛАДНИХ МЕРЕЖАХ.....</b>	138
<i>Subach I.Y., Kubrak V.O., Mykytiuk A.V.</i> <b>METHODOLOGY OF RATIONAL CHOICE OF SECURITY INCIDENT MANAGEMENT SYSTEM FOR BUILDING OPERATIONAL SECURITY CENTER.....</b>	146
<i>Tmienova N., Sus' B.</i> <b>SYSTEM OF INTELLECTUAL UKRAINIAN LANGUAGE PROCESSING.....</b>	152

<i>Tokariiev V. , Tkachov V. , Ilina I., Stanislav P.</i>	
<b>IMPLEMENTATION OF COMBINED METHOD IN CONSTRUCTING A TRAJECTORY FOR STRUCTURE RECONFIGURATION OF A COMPUTER SYSTEM WITH RECONSTRUCTIBLE STRUCTURE AND PROGRAMMABLE LOGIC .....</b>	<b>159</b>
<i>Yartsev V., Hololobov D.</i>	
<b>PROTECTION DATA TRANSMISSION SYSTEMS FROM THE INFLUENCE INTERSYMBOL INTERFERENCE SIGNALS.....</b>	<b>166</b>
<i>Юзефович В.</i>	
<b>МОДИФІКОВАНИЙ МЕТОД ЕКСПОНЕНЦІАЛЬНОГО ЗГЛАДЖУВАННЯ ДЛЯ ФІЛЬТРАЦІЇ КУРСУ РУХОМИХ ОБ'ЄКТІВ ПРИ ЇХ СУПРОВОДЖЕННІ.....</b>	<b>173</b>
<i>Гнатієнко Г.М.</i>	
<b>МАНІПУЛЮВАННЯ ВИБОРОМ У ЗАДАЧАХ БАГАТОКРИТЕРІАЛЬНОЇ ОПТИМІЗАЦІЇ</b>	<b>179</b>
<i>Гнатієнко Г.М., Снитюк В.С.</i>	
<b>АПОСТЕРІОРНЕ ВИЗНАЧЕННЯ КОМПЕТЕНТНОСТІ ЕКСПЕРТІВ В УМОВАХ НЕВИЗНАЧЕНОСТІ.....</b>	<b>184</b>
<i>Додонов О.Г., Кузьмичов А.І.</i>	
<b>ЖИВУЧІСТЬ Й КОМПРОМІС: ФОРМУВАННЯ ПАРЕТО-ОПТИМАЛЬНИХ РІШЕНЬ ОРГАНІЗАЦІЙНОГО УПРАВЛІННЯ ЗАСОБАМИ EXCEL.....</b>	<b>188</b>
<i>Зубок В.Ю.</i>	
<b>ПОБУДОВА ФОРМАЛЬНОЇ МОДЕЛІ ІНТЕРНЕТ-МАРШРУТИЗАЦІЇ ДЛЯ ОЦІНКИ ВПЛИВУ АТАК З ПЕРЕХОПЛЕННЯМ МАРШРУТІВ.....</b>	<b>196</b>
<i>Ланде Д.В., Боярінова Ю.С., Каліновський Я.О., Синькова Т.В.</i>	
<b>ЗАСТОСУВАННЯ ГІПЕРКОМПЛЕКСНИХ ЧИСЛОВИХ СИСТЕМ ДЛЯ ОПИСУ СКЛАДНИХ МЕРЕЖ.....</b>	<b>201</b>
<i>Матов О. Я.</i>	
<b>АДАПТАЦІЯ ХМАРНИХ ОБЧИСЛЕНЬ ЯК ОПТИМІЗАЦІЯ ПРОЦЕСУ НАДАННЯ ПОСЛУГ КОРИСТУВАЧАМ В УМОВАХ ОБМЕЖЕНИХ ОБЧИСЛЮВАЛЬНИХ РЕСУРСІВ</b>	<b>210</b>
<i>Савченко М.М., Циганок В.В., Андрійчук О.В.</i>	
<b>ПІДХІД ДО ДЕЛЕГУВАННЯ ТРАНЗАКЦІЙ У САМОЗАХИСНИХ ДЕЦЕНТРАЛІЗОВАНИХ ПЛАТФОРМАХ ДАНИХ.....</b>	<b>215</b>
<i>Цуркан О., Герасимов Р., Крук О.</i>	
<b>СПОСОБИ ПРОТИДІЇ ВИКОРИСТАННЮ СОЦІАЛЬНОЇ ІНЖЕНЕРІЇ.....</b>	<b>229</b>
<i>Додонов О.Г., Ланде Д.В., Нестеренко О.В., Березін Б.О.</i>	
<b>ПІДХІД ДО ПРОГНОЗУВАННЯ ДІЄВОСТІ ДЕРЖАВНОГО УПРАВЛІННЯ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ OSINT.....</b>	<b>230</b>