

РОМАН КАМІНСЬКИЙ, НАТАЛІЯ ШАХОВСЬКА, БОГДАН ХУДОБА
Національний університет «Львівська політехніка»

ФРАКТАЛЬНИЙ АНАЛІЗ МОДЕЛЕЙ ТЕКСТІВ РІЗНИХ СТИЛІВ, ПОДАНИХ ЦІЛОЧИСЕЛЬНИМИ ЕКВІДИСТАНТНИМИ ПОСЛІДОВНОСТЯМИ КІЛЬКОСТІ ЛІТЕР У СЛОВАХ

В даній роботі приведені результати фрактального аналізу моделей текстів різних стилів. В якості моделей використано цілочисельні числові послідовності, елементами яких є кількість літер у словах тексту. Наведено алгоритм обчислення точного значення фрактальної розмірності, який забезпечив визначення точного значення показника Герста. Крім того, обчислено значення константи степеневі залежності R/S . Отримані показники в аспекті фрактальності повністю описують об'єкти дослідження.

Ключові слова: моделювання тексту, цілочисельні дані, фрактальна розмірність, клітинковий метод, показник Герста, константа степеневі функції.

ROMA KAMINSKYI, NATALIYA SHAKHOVSKA, BOHDAN KHUDOBA
Lviv Polytechnic National University

FRactal analysis of models of texts of different styles submitted integer equidistant sequences number of letters in words

This paper presents the results of fractal analysis of models of texts of different styles. Integer numerical sequences, the elements of which are the number of letters in the words of the text, are used as models. An algorithm for calculating the exact value of the fractal dimension is presented, which provided the determination of the exact value of the Hirst index. In addition, the value of the power dependence constant R/S is calculated. The obtained indicators in the aspect of fractality fully describe the objects of research. This method is in fact a logical implementation of the known procedures of fractal analysis and its advantage is that it provides a rigorous mathematical representation of the values of the fractal dimension, the Hirst index and the constant in relation to the indicators of variation. The essence of his presentation is, first of all, as a warning to researchers against misinterpretation of the relationship R/S , because many researchers ignore the existence of a constant for this relationship. Indeed, this relation is a function with two unknown parameters and cannot be directly determined. With regard to the fractal dimension, we can point out that the least important is the conversational style, and the most - poetic. In other words, the model of colloquial text is the smallest part of its environment, poetic - the largest. From the point of view of Hirst's index, the model of the spoken text contains a trend, while the model of the poem has a character closer to the random one. The largest scope of the cumulative series has a model of spoken text, and the smallest - a model of artistic style of the text. Since the cumulative series is a sequential (cumulative) summation of the sequence of deviations of elements from its arithmetic mean, its scope will depend on the presence of groups of elements of the sequence with very large deviations. Artistic style has the least significance of scope.

Keywords: text modeling, integer data, fractal dimension, cellular method, Hirst index, power function constant.

Вступ

Однією з актуальних задач лінгвістики в аспекті розробки систем аналізу текстів є розгляд їх з точки зору математичних об'єктів. В цьому сенсі досліджуваний текст має бути представлений відповідною математичною моделлю. Виходячи з структури тексту як послідовності речень і речень як послідовності слів, в якості такої моделі тексту авторами вибрана цілочисельна еквідистантна послідовність.

Суть моделі в тому, що послідовності слів тексту замінюють числами, кожне з яких відповідає кількості літер у словах цього тексту. Очевидно, що допоміжні розділові знаки початково можуть бути вилучені, а саму послідовність можна прийняти за регулярну. Така послідовність є аналогічною до еквідистантного часового ряду зі значеннями рівнів, поданих цілими числами. У випадку тексту така послідовність не зв'язана з часом, як часовий ряд, а підпорядкована послідовності слів. Тим не менш, вона є математичним об'єктом і до неї можуть бути застосовані такі ж самі методи як і до часових рядів.

Останнім часом, а точніше за останніх два десятиліття, стосовно часових рядів, вельми загальноприйнятним виявився фрактальний підхід та його методологія – фрактальний аналіз. Фрактальний аналіз на сьогодні присутній практично в будь-якій області, яка пов'язана з обробкою різних даних, оскільки забезпечує використання декількох, досить важливих характеристик для їх класифікації, інтерпретації та прогнозування. До таких характеристик відносять фрактальну розмірність часового ряду, показник трендовості часового ряду Герста, константу R/S відношення зв'язку між показниками варіації – розмахом кумулятивного ряду та середньоквадратичного значення цього ж ряду.

У цьому дослідженні не йде мова про фрактальний аналіз власне самого тексту, а лише стосується застосування фрактального аналізу до саме такої моделі для різних текстів та виявити між ними відмінності у фрактальних показниках. Власне, тут робоча гіпотеза є така: чи може, за результатами фрактального аналізу виявити відмінності серед таких моделей текстів. Тому, *метою* даного дослідження є визначення правомірності застосування фрактального аналізу для моделі тексту у вигляді цілочисельної послідовності кількості літер у словах тексту, за відсутності усіх додаткових знаків.

Аналіз публікацій дотичних до даного дослідження.

Моделювання текстів, зокрема математичним апаратом не є чимось новим. Модель у вигляді регулярної цілочисельної послідовності може бути віднесена до класу потоку подій лише з точки зору випадкового числа літер у словах. Тут саме кількість літер визначає випадковість, а не ймовірність появи

конкретного слова. В даному випадку, зразу ж виникає питання: який зв'язок існує між послідовністю слів у реченнях тексту і кількістю літер у словах цих речень. На жаль, в лінгвістичній літературі, ні відповіді про існування такої моделі тексту, ні, тим більш, чи існує якийсь зв'язок між кількістю літер в словах і їх змістом, авторами не виявлено, проте переконав у широкому застосуванні фрактального аналізу.

В [1] приведена методика R/S аналізу з поданням показника Герста, який враховує роль константи C . Визначення фрактальної розмірності різними методами приведені в [2], а саме: з використанням показника Герста H та за допомогою кореляційного інтегралу, яке відоме як метод або алгоритм Грассбергера-Прокаччіа. Оцінювання експоненти Герста в [3] здійснено трьома методами: R/S, DFA та вейвлет аналізу, а також приведені результати порівняння цих методів. В [4] приведені кілька зауважень стосовно фрактальної розмірності та показника Герста.

В [5] приведено методи оцінювання фрактальної розмірності, показника Герста і що вельми важливо, приведений аналітичний спосіб обчислення константи степеневі залежності відношення R/S. Два методи визначення – R/S аналізу та сегментово-варіаційного методу представлено в [6]. Дана сегментово-варіаційного методика досить близька методу. Який використовують автори даного дослідження. В роботі [7] представлений взаємозв'язок між показником Герста та R/S аналізом стосовно класифікації часових рядів валютного ринку. Показано, що показник Герста є метрикою, здатною надати інформацію про кореляцію та стійкість у часовому ряді. Книга [8] дає ясний, доступний і простий виклад математичних властивостей фрактальних об'єктів та часових рядів, зокрема фрактальної розмірності та показника Герста. В [9] розкриваються особливості клітинкового методу визначення фрактальної розмірності, зокрема вказується, що в підрахунок включають і частину неповністю заповнених клітинок. Крім того, зростає складність підрахунку клітинок, коли їх розмір зменшувати. В [10] проаналізовано вплив зовнішнього адитивного шуму на клітинковий алгоритм розрахунку фрактальних показників часових рядів скінченних обсягів. Було виявлено, що ефект шуму напрочуд великий – відносно незначний зовнішній шум передбачає зростання значення похибки на три-чотири порядки більше. Невеликий шум, як правило, є частиною будь-яких реальних досліджуваних даних. Як зазначено в цій статті, слід бути дуже обережним, роблячи висновки на основі чисельно вираховуваних фрактальних показників для експериментальних даних. Основні типи моделей, які застосовуються в лінгвістичних дослідженнях та їх використання для розв'язку різних лінгвістичних задач розглянуто в [11]. Тут окреслено основні підходи щодо розуміння поняття моделі у лінгвістиці. В [12] розглядаються актуальні для математичного моделювання лінгвістичних об'єктів підходи, обґрунтовується доцільність застосування математичних методів, приведені основні принципи створення математичних моделей. Для усунення недоліків існуючих моделей текстових документів, в [13] запропоновано уніфіковану форму змістовної моделі тексту, в основі якої лежить синтез логіко-лінгвістичних моделей його речень, а також описано алгоритм побудови такої моделі. В [14] досліджено теоретичні питання використання моделювання в лінгвістиці, а також зроблено акцент на лінгвістичних моделях, їх особливостях. Також, в лінгвістичному аспекті, описані характерні ознаки моделей та основні етапи їх створення. Крім того, вказано основні галузі, де метод моделювання якісно змінив парадигму лінгвістичного дослідження. Елементи теорії та застосування цілочисельних потоків висвітлені в [15]. Вказано, що такі потоки є елементарними еквідистантними потоками подій з випадковими значеннями амплітуд.

Про фрактальні властивості тематичних інформаційних потоків з Інтернету йдеться в [16], причому, в якості бази даних для обчислювального експерименту вибрана система моніторингу мережевих новин InfoStream. Представлена методика обчислення показників Герста для кластера, визначеного тематикою запиту та приведена якісна інтерпретація результатів. В [17] показано, аналіз інформаційних потоків став одним з основних методів пошуку закономірностей функціонування світової системи наукової комунікації. Основи інтеграції інформаційних потоків висвітлено в [18], де також приведені математичні моделі, елементи теорії інформаційного пошуку та концепції глибинного аналізу тексту (Text Mining) до інформаційних потоків, В [19] розглянуто та подано характеристики основних напрямів та методів та організації лінгвістичних досліджень. До одного із важливих напрямів квантитативних досліджень мови та мовлення належить праця [20], в якій розглянуто вивчення інформаційно-статистичних властивостей тексту. Приведені розрахунки довжину речення та слова у творах Р. Іваничука. Отримані результати зіставлено з аналогічними показниками в українській прозі.

1. Поняття та моделі регулярної цілочисельної послідовності

Аналіз будь-яких зібраних для вивчення даних, зокрема лінгвістичних, має дати відповідь щодо особливостей і властивостей джерела, що породжує ці дані. Іншими словами, повнота отриманих показників фактично вказує на особливості такого джерела, проте ніяк не підказує його фізичну природу, а в обробці текстів автора тексту.

Одним з видів експериментальних даних є потік подій, які є випадковими величинами зі своїм законом розподілу. Такий потік дискретних випадкових величин в літературних джерелах та дослідниками прийнято називати часовим рядом. Особливістю часових рядів є те, що значення його елементів (рівнів) відповідають фіксованим, тобто конкретним моментам часу. Наприклад кількість виробів з наданого обсягу ресурсу, кількість бракованих деталей в послідовності випущених партій деталей тощо.

Моменти фіксації рівнів можуть бути як регулярними – еквідистантний часовий ряд, тобто фіксація відбувається через рівні проміжки часу, так і нерегулярними – випадковими, коли рівні фіксують у

випадкові моменти часу. Якщо мова не йде про фіксацію рівнів в часі, такий ряд називають числовою послідовністю. Одним з різновидів такої послідовності є послідовність з цілочисельними значеннями величин рівнів. Оскільки елементи послідовності приймають цілі додатні значення, на графіку ці величини будуть відповідати поділкам шкали ординат. Шкала ординат в цьому випадку відповідає величині значення кількості літер у словах, а шкала абсцис – порядку слідування цих слів.

Візуально, характер подання таких дискретних послідовностей може бути різним. Наприклад, у вигляді діаграми або як регулярну послідовність імпульсів однакової тривалості, але різної (переважно випадкової) амплітуди як на рис. 1 зліва. Якщо значення елементів подати точками, причому дві сусідні точки сполучені відрізками ліній, то маємо звичайне графічне зображення випадкового процесу, як на рис. 1 справа.

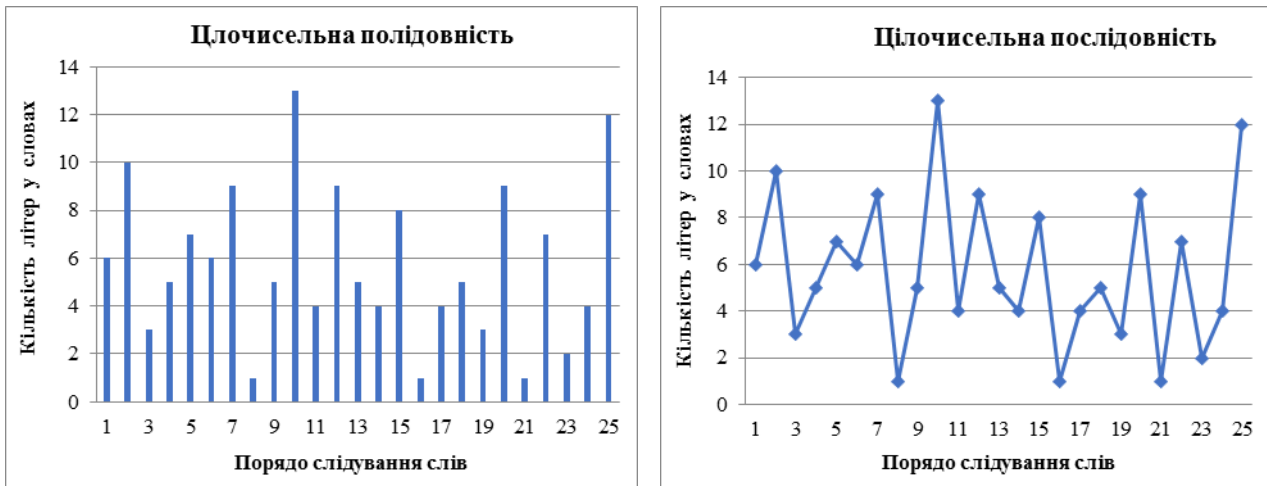


Рис. 1. Види представлення числових цілочисельних послідовностей: зліва діаграмою, справа графіком випадкового процесу

В більшості випадків таку сукупність елементів можна розглядати як елементи – варіанти звичайної вибірки, проте, якщо події настають одна за одною регулярно чи випадково маємо справу з дискретним процесом.

Очевидно що така послідовність може бути представлена як:

- регулярна послідовність елементів, коли інтервали між ними є точно заданими і не змінюють свого розміру:

$$z_{i+1} - z_i = \delta = const ;$$

- мають місце допустимі відхилення від регулярності, тоді

$$z_{i+1} - z_i = \delta \pm \Delta = var ;$$

- інтервали є випадковими величинами

$$z_{i+1} - z_i = \rho = random .$$

В такому поданні послідовність представляє аналог дискретного процесу бо відстань між елементами є фактично тривалістю до появи наступного елемента.

Дискретність послідовності елементів матиме функціональне представлення якщо значення елементів є числовими величинами. В даному випадку розглядаються цілочисельні додатні значення, тобто послідовність рівнів має такий вигляд

$$z_i \in Z = \{ z_i | z_1, z_2, \dots, z_i, \dots, z_{n-1}, z_n, i \in I, \} , \tag{1}$$

причому множини $Z, I \subset \mathbf{Z}^+$, де Z – множина значень елементів z_i , а I – множина їх індексів.

Формування моделі виконано за допомогою табличного процесора Ms Excel.

Фактично маємо послідовність випадкових подій, задану конкретним генератором. Ці події можуть бути одновимірними, тобто характеризувати одну властивість або бути багатовимірними і характеризуватися вектором відповідних ознак (наприклад частота вживання за словником, кількість синонімів, ступінь відношення до тематики тощо), що описують їх властивості, особливості та зв'язки з текстом.

2. Експериментальне дослідження словесних потоків

2.1. Підготовка текстів до аналізу. В якості об'єктів дослідження використано наведені нижче тексти різних стилів, а також вказані адреси їх джерел. Крім того, вказано обсяги кількості слів їх семантичної повноти, тобто повного тексту.

Кожен, використаний в експерименті текст відповідає конкретному стилю. Правомірність вибору саме цих текстів, зумовлена тим, що вони представлені на вказаних сайтах та публікаціях як ілюстрація даного стилю. Отже, вибраними стилями є такі:

- **розмовний** (<https://ycilka.net/tvir.php?id=291>), обсягом слів 259;

- *художній* стиль (<http://litmisto.org.ua/?p=5462>), 152 слова;
- *науковий* стиль (<https://ridna-mova.com.ua/engine/download.php?id=177>), слів 215;
- *офіційно-діловий* стиль [21], 179 слів;
- *публіцистичний* стиль (<https://www.ukrlib.com.ua/sochm/printout.php?id=944>), 229 слів;
- *конфесійний* стиль, 208 слів (<http://news.ugcc.ua/bible-quote/%D0%94%D1%96%201:1-8,%20%D0%99%D0%BE%201:1-17>)
- *епістолярний* стиль, 223 слова, <https://sites.google.com/site/stilistikamiller/home/epistolarnij-stil>;
- *віри* (Л. Костенко «І все на світі треба пережити» – <https://prm.ua/51198-14/>).

Подання цих текстів моделями у вигляді цілочисельних послідовностей є досліджуваними математичними об'єктами. Крім того, ці моделі текстів подаються як еквідистантні послідовності цілочисельних елементів. Представлені в них тексти це лише послідовність елементів, значення яких представлених кількістю літер у словах, що знаходяться в даній послідовності на їх місці цих елементів. Побудова таких моделей полягає в тому, що вибрані для дослідження тексти перетворюють в числові послідовності в такий спосіб.

З текстів вилучено усі розділові та будь-які інші знаки, включаючи апостроф і дефіс. Розділювачем між словами залишається пробіл. Оскільки вибрані уривки текстів склалися з різного обсягу слів, було вирішено обмежити їхні обсяги до однакової величини, а саме до обсягу 150 слів. Справа в тому, що для фрактального аналізу обсяг елементів послідовності є незалежною *змінною* і його величина впливає на обчислювані показники. В даному випадку йде мова про процедуру фрактального аналізу, а тому при його проведенні для кількох різних текстів необхідно, щоб усі вони були в однакових умовах. Це у свою чергу підвищує значення висновків щодо застосування фрактального аналізу, особливо в задачах ідентифікації текстів.

Такий «очищений» текст записується в таблицю так, кожне слово було в окремій комірці, причому має бути збережена послідовність слів в напрямку зверху вниз. На рис. 2 зображено графічне подання моделі тексту розмовного стилю як послідовності слів і кількості літер у словах.



Рис. 2. Графічна модель тексту розмовного стилю

Далі визначається кількість літер у кожному слові.

У результаті буде отримана таблиця з двох стовпців: лівий для слів, а правий для кількості літер в них.

2.2. Формальне подання клітинкового методу. Мінімальний розмір клітинок сітки означає те, що для еквідистантної числової послідовності значення кожного її елемента має свою одну і єдину клітинку. Виходячи з того, що клітинки сітки мають форму квадрату, їх розмір строго визначений поділками шкали ординат. В цьому випадку, накладання сітки мінімального на графік даної послідовності означатиме, що значення будь якого елемента, відображене відповідними координатами $z(x_i, y_i)$, буде локалізоване в межах однієї клітинки і кількість зайнятих клітинок буде рівна кількості елементів цієї послідовності.

Для визначення кількості клітинок конкретної сітки необхідно встановити їх мінімальний розмір та кратність його зміни. Справа в тому, що за умови регулярності послідовності кожна сітка має відповідати дискретності послідовності, тобто розміри клітинок сітки є кратні до розмірів клітинки для одного елемента. На рис. 2 зображено (чорними лініями), які є кратні до мінімальної клітинки (зображені сірим кольором), що вміщує один елемент послідовності з мінімальним значення кількості літер одну, наприклад, слово № 8. Щоб визначити кількість клітинок сітки даного розміру покривають графік послідовності розбивають послідовність на групи по два елемента, по три, по чотири, по п'ять і по шість.

На цьому рисунку чітко видно як локалізовані маркери елементів послідовності в залежності від розмірів клітинок сіток. Варто звернути увагу на те, що для цілочисельних значень величин елементів положення центрів маркерів збігається з поділками шкали кількості літер. Для послідовності цілочисельних величин це забезпечує точний підрахунок клітинок.

Отже, за умови, що послідовність є еквідистантною, підрахунок кількості клітинок автори пропонують здійснювати, за допомогою табличного процесора, в такий спосіб.

Для кожної групи визначають кількість клітинок сітки, що охоплюють по вертикалі клітинки з мінімальним і максимальним значеннями елементів в кожній групі. Тут мається на увазі такий підрахунок: для груп по два елементи розмір клітинок сітки є 2×2 поділкі шкали на осі ординат для значень елементів, для групи по три елементи відповідно 3×3 поділкі цієї шкали, аналогічно для решти груп.

Фрактальну розмірність D розраховують так (рис. 3).

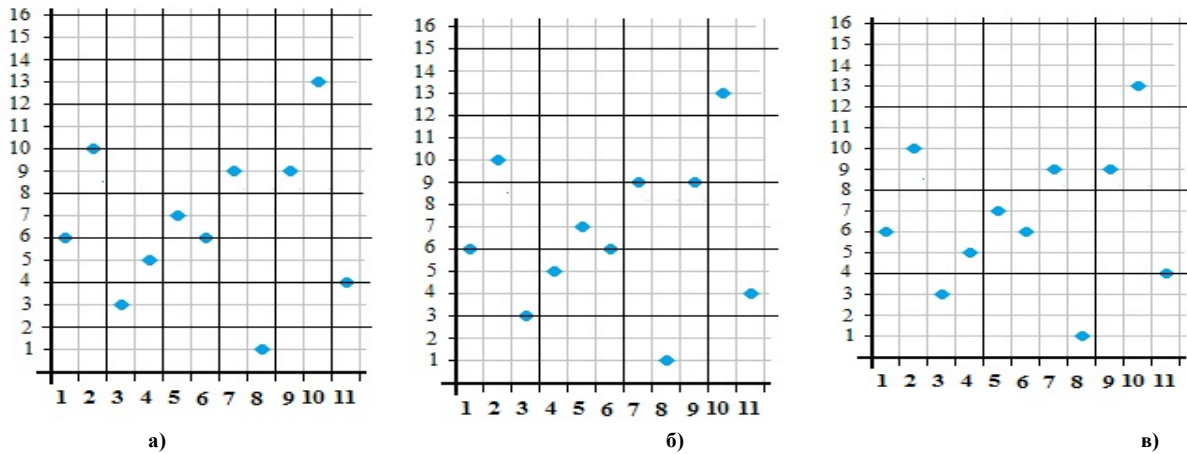


Рис. 3. Набір сіток накладених на графічне зображення послідовності. Тут маркери графіка вказують кількість літер в послідовності слів

1) для кожної групи і для кожної сітки обчислюють між максимальним і мінімальним значеннями елементів, що входять в цю групу і отриману величину ділять на розмір клітинки сітки. Очевидно, що розмір групи і розмір клітинки мають збігатися, тобто кількість клітинок q_m конкретного розміру сітки може бути визначена так:

$$q_m = \sum_{j=1}^p \left(\frac{\max(z_p^m) - \min(z_p^m)}{d_m} \right), \quad (5)$$

де z_p^m – елемент послідовності, m – номер (розмір клітинок) сітки, p – кількість груп розбиття послідовності під конкретну сітку, d_m – розмір клітинки даної сітки.

2) У результаті сумування значень величин, що є в дужках, буде отримана кількість клітинок q_m . Значення q_m є загалом дробовим числом, проте воно є точним значенням кількості клітинок даного розміру які покривають дану послідовність.

2.3. Визначення фрактальної розмірності D послідовності. Фрактальну розмірність обчислюють клітинковим методом Гаусдорфа-Безіковича. Цей метод має два суттєвих недоліки. Першим є те, що при безпосередньому використанні він дає завищені значення кількості клітинок, а отже і вносить відповідну похибку в розрахунки. Другим недоліком цього методу є те, що підрахунок кількості клітинок, особливо для довгих рядів потребує розробки спеціальних програм, що вимагає від дослідників вміння програмувати або принаймні знайти відповідні програми та навчитися їх використовувати.

Авторами запропоновано і реалізовано в даному дослідженні **модифікований метод обчислення фрактальної розмірності** з визначенням співвідношення заповненості клітинок.

З першої формули системи (4) впливає алгоритм обчислення D , суть якого полягає ось в чому. Фрактальна розмірність є параметром функції внутрішнього зв'язку, яка характеризує залежність відношення

$$D = \lim_{\delta^m \rightarrow 0} \frac{\log(n^m)}{\log(\delta^m)},$$

тобто, відношення логарифму кількості клітинок сітки (з номером m), що покривають до логарифму розміру δ^m клітинок цієї сітки. Значення цього параметру визначають як абсолютне значення кутового коефіцієнта апроксимуючої точки графіка лінії в подвійній логарифмічній системі координат, які відповідають: абсцисі – $\log(\delta)$ і ординаті – $\log(N)$. Як показано було вище для визначення фрактальної

розмірності необхідно подати зв'язок між кількістю клітинок сітки, що покривають графік послідовності і розміром клітинок даної сітки

N	O	P	Q	R	S	T
Номер слова	К-сть літер	m=2	m=3	m=4	m=5	m=6
1	6	2	2,333	1,75	1,4	1,167
2	10	3,5	2,333	1,75	1,4	1,167
3	3	1	1,333	1	1,2	1,333
4	5	1	0,667	1	1,6	1,333
5	7	0,5	1	2	1,6	2
6	6	1,5	2,667	2	2,4	2
7	9	4	2,667	3	2,4	2
8	1	2	4	3	2,4	2
9	5	4	3	2,25	1,8	1,5
10	13	4,5	3	2,25	1,8	1,5
11	4	0	0	0	0	0

Рис. 4. Організація обчислення кількості клітинок використаних сіток

Будуємо графік в подвійному логарифмічному масштабі за приведеними значеннями логарифмів, як зображено на рис. 5.

Значення фрактальної розмірності D визначається абсолютною величиною кутового коефіцієнта для рівняння лінії тренду графіка, тобто фрактальна розмірність даної послідовності дискретних величин у випадку даного тексту є рівна $D = 1,3105$.

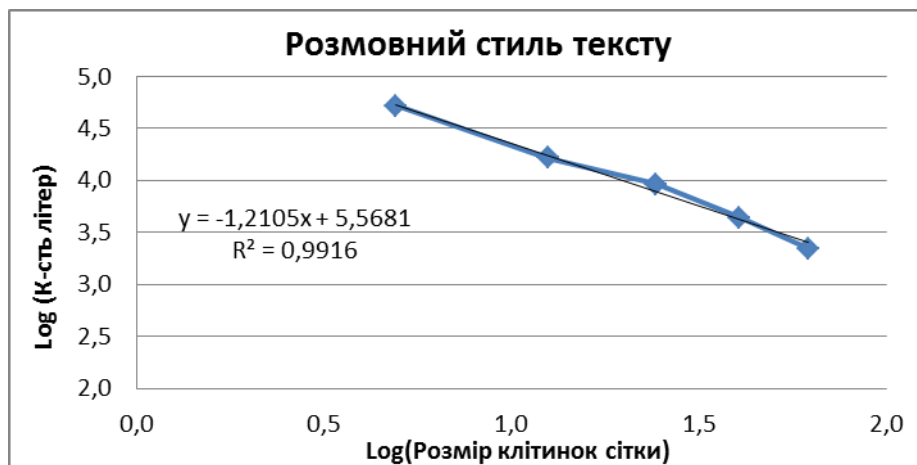


Рис. 5. Апроксимація фрактальних значень

2.4. В значення експоненти Герста. Показник Герста характеризує тип процесу, який домінує в динаміці часового або числового ряду.

На сьогодні значення показника Герста визначають різними методами, переважно використовуючи рівняння (3) нехтуючи константою C , що свідчить про відсутність знань елементарної математики. Цей показник легко визначити з другого рівняння, включеного в систему (4), тобто з рівності вказаної Б. Мандельбротом, а саме:

$$H = 2 - D.$$

Іншими словами, на відміну від описаного підходу в переважаючій більшості відповідних публікацій, значення експоненти H визначається через значення фрактальної розмірності D .

Отже, використовуючи формулу (2) значення показника Герста є рівне $H = 0,7895$.

Показник Герста показує, якою є частка присутнього в ряді хаосу. Як відомо, таких типів поведінки може бути три: коливання значень елементів ряду, випадкова зміна значень рівнів, близька до броунівського руху та існування тренду.

Область визначення показника Герста – відрізок $[0, 1]$, а його зміст заключений в таких межах:

якщо $0 \leq H < 0,5$ – рівні мають коливальний характер;

якщо $H = 0,5$ – ряд є прикладом випадкового броунівського руху;

якщо $0,5 < H \leq 1$ – ряд є фрактальним з наявністю тренду;

2.5. Визначення константи R/S відношення.

Відношення R/S є відношенням двох показників варіації, зв'язаних між собою середнім

арифметичним. Відносна зміна однієї з цих величин зумовлює пропорційну відносну зміну іншої величини, незалежно від їх початкового розміру. За своєю природою це відношення є степеневу функцією і формально описується третім рівнянням системи (4), з якого, за відомого значення H , вираз для значення константи має такий вигляд:

$$C = \frac{R_n}{S_n \cdot n^H}$$

Тут R_n є розмахом кумулятивного ряду; S_n – середньоквадратичне відхилення; C – константа; m – кількість даних; H – експонента Герста. Константа C є фактично коефіцієнтом пропорційності і характеризує природу генератора будь-яких даних: чи то числової послідовності чи часового ряду.

Обговорення

В результаті проведення відповідних обчислень над моделями текстів різних стилів отримані такі результати фрактального аналізу, приведені в табл. 1.

Таблиця 1

Результати фрактального аналізу

Стилі	Параметри					
	n	D	H	R	S	C
розмовний	150	1,2105	0,7895	47,4362	2,742	0,3329
художній	150	1,3497	0,6503	23,4228	2,3791	0,3802
науковий	150	1,3157	0,6843	36,9329	3,0744	0,3913
діловий	150	1,2607	0,7393	33,7315	3,2267	0,2586
публіцистичний	150	1,2414	0,7586	32,3893	2,9368	0,2477
конфесійний	150	1,3485	0,6515	27,7248	2,3301	0,4567
епістолярний	150	1,3129	0,6871	41,3557	3,2989	0,4027
віршований	150	1,3702	0,6298	37,1611	2,3408	0,6793

Стосовно фрактальної розмірності можна вказати на те, що найменше її значення має розмовний стиль, а найбільше – віршований. Іншими словами, модель розмовного текст найменшу частину свого середовища, віршований – найбільшу. З точки зору показника Герста модель розмовного тексту містить тренд, той час як модель вірша має характер ближчий до випадкового.

Найбільший розмах кумулятивного ряду має модель розмовного тексту, а найменший – модель художнього стилю тексту. Оскільки кумулятивний ряд це послідовне (накопичувальне) сумування послідовності відхилень елементів від свого середнього арифметичного, його розмах буде залежати від присутності груп елементів послідовності з дуже великими відхиленнями. Найменше значення розмаху має художній стиль. Найбільшу величину середньоквадратичного відхилення має епістолярний стиль. Найменше значення константи степеневу функції R/S має публіцистичний стиль, а найбільше значення цієї константи має віршований. Значення цієї константи відрізняється майже в три рази, і це лише для восьми коротеньких текстів.

Звичайно, залишається відкритим питання як співвідносяться між собою оригінальний текст і його модель у вигляді цілочисельної, еквідистантної послідовності, яким є зв'язок між словами тексту і кількістю літер в цих слова. Очевидно, що на ці питання, в сенсі позитивної чи негативної відповіді, можуть дати лише подальші дослідження. Приведені результати, в жодному випадку не можна вважати істиною в останній інстанції, оскільки вони є лише першою спробою перевірки гіпотези – щодо наукової і практичної цінності такої моделі тексту в задачах комп'ютерної лінгвістики та штучного інтелекту стосовно аналізу текстів.

Висновок

Проведене дослідження за запропонованою методикою дало підстави для її практичного використання. Проте, таке дослідження вимагає обов'язкової присутності висококваліфікованого фахівця – лінгвіста в області стилеметрії.

Можна по-різному відноситися до фрактального аналізу числових послідовностей – вважати його статистичним методом чи одним з методів нелінійної динаміки, а ще вважати його окремою методологією. Крім того, при огляді публікацій розуміння авторами суті цих методів, фактично лише двох основних, до яких відносять фрактальну розмірність D та показник Герста H . Усі інші методи впливають з показника Герста.

Приведений метод є фактично логічною реалізацією відомих процедур фрактального аналізу і його перевага в тому, забезпечує строге математичне подання значень фрактальної розмірності, показника Герста та константи у відношенні показників варіації. Суть його подання полягає, в першу чергу, як застереження дослідників від невірного трактування відношення R/S , оскільки досить багато дослідників ігнорують існування константи для цього відношення. Дійсно, це відношення є функцією з двома невідомими параметрами і безпосередньо визначити їх неможна.

Література

1. Steinhaus H. Zacharewicz A. Metody analizy długozasiękowej. 2002, 213 p.
2. Кириллов А.К. Фрактальная размерность временных рядов биологических процессов и объектов / А.К. Кириллов, А.Н. Казимиров // Вісник Донецького національного університету, Сер. А: Природничі науки. – 2013. – № 2. – С. 149–155.
3. J. Mielniczuk, P. Wojdyło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), DOI:10.1016/j.csda.2006.07.033.
4. Kowgier H. Kilka uwag o wymiarze fraktalnym Minkowskiego oraz wykładniku Hursta na giełdzie papierów wartościowych. Studia i prace wydziału nauk ekonomicznych i zarządzania. N 15, 2018, p. 157–167.
5. Kowgier H. Kilka uwag o wymiarze fraktalnym Minkowskiego oraz wykładniku Hursta na giełdzie papierów wartościowych. Studia i prace wydziału nauk ekonomicznych i zarządzania NR 15, 2016, str. 157–167.
6. Orzeszko W. Wymiar fraktalny szeregów czasowych a ryzyko inwestowania / Acta Universitatis Nicolai Copernici Ekonomia Xli – Nauki Humanistyczno-Społeczne – Zeszyt 397, Toruń 2010, P. 57–70.
7. Milton S. Raimundo. Application of Hurst Exponent (H) and the R/S Analysis in the Classification of FOREX Securities. International Journal of Modeling and Optimization 8(2): 116-124 DOI: 10.7763/IJMO.2018.V8.635
8. Федер Е. Фракталы / Федер Е. ; пер. с англ. – М. : Мир, 1991. – 254 с.
9. Zwolankowska M. Metoda segmentowo-wariacyjna. Nowa propozycja liczenia wymiaru fraktalnego // Przegląd Statystyczny, 2000, R. 47, z. 1-2, S. 209–224.
10. Skrzat J. Accuracy of the box-counting algorithm for noisy fractals. International Journal of Modern Physics C December 2014. DOI: 10.1142/S0129183116501126 – <https://www.researchgate.net/publication/269935552>
11. Лукач М.О. Типи лінгвістичних моделей та їх застосування для розв'язання лінгвістичних задач, 2016. – URL : http://science.lp.edu.ua/sites/default/files/Papers/19_151.pdf
12. Васильєв О. Математичні методи й моделі в лінгвістиці / О. Васильєв, О. Чалий, І. Васильєва. – URL : https://uamoderna.com/images/archiv/27-2020/27_9_28%20Oleksii%20VASYLIEV_compressed.pdf
13. Вавіленкова А. Побудова змістовної моделі тексту на основі використання логіко-лінгвістичних моделей. – Lviv Polytechnic National University, 2016. – С. 169–175. – URL : <http://ena.lp.edu.ua:8080/bitstream/ntb/31313/1/25-169-175.pdf>
14. Калимон Ю.О. Моделирование: лингвистический аспект / Ю.О. Калимон // Молодой ученый. – 2019. – № 4.2 (68.2). – С. 88–91.
15. Большаков И.А. Прикладная теория случайных потоков / Большаков И.А., Ракошиц В.С. – М. : Сов. Радио, 1978. – 248 с.
16. Ландэ Д.В. Фрактальные свойства тематических информационных потоков из Интернет / Д.В. Ландэ // Реєстрація, зберігання і обробка даних. – 2006. – Т. 8, № 2. – С. 93–99.
17. Иванов С.А. Статистический анализ документальных информационных потоков / С.А. Иванов, Н.В. Круковская // НТИ. СЕР. 2. ИНФОРМ. ПРОЦЕССЫ И СИСТЕМЫ. – 2004. – № 2. – С. 11–14.
18. Ландэ Д.В. Основы интеграции информационных потоков : монография / Ландэ Д.В. – К. : Инжиниринг, 2006. – 240 с.
19. Богдан С. К. Методи й методика лінгвостилістичних досліджень: методичні рекомендації для слухачів і керівників секції української мови / С. К. Богдан. – Луцьк, 2011. – 28 с.
20. Кульчицький І.М. Дослідження довжини речення та слова у творах Романа Іванчука. – URL : <http://science.lpnu.ua/sites/default/files/journal-paper/2018/jun/13011/ilovepdfcom-139-148.pdf>
21. Заверющенко М.П. Офіційно-діловий стиль: правила укладання документів різних видів : навч. посіб. щодо самостійної роботи студентів / М.П. Заверющенко, О.М. Кривець, С.М. Чернявська, О.В. Шокуров. – Харків : НТУ «ХПІ», 2019. – С. 93.

References

1. Steinhaus H. Zacharewicz A. Metody analizy długozasiękowej. 2002, 213 p.
2. Kirillov A.K. Fraktal'naya razmernost' vremennykh ryadov biologicheskikh processov i ob'ektov / A.K. Kirillov, A.N. Kazimirov // Visnik Donec'kogo nacional'nogo universitetu, Ser. A: Prirodnichi nauki. – 2013. – № 2. – S. 149–155.
3. J. Mielniczuk, P. Wojdyło, Estimation of Hurst exponent revisited, Computational Statistics & Data Analysis (2006), DOI:10.1016/j.csda.2006.07.033.
4. Kowgier H. Kilka uwag o wymiarze fraktalnym Minkowskiego oraz wykładniku Hursta na giełdzie papierów wartościowych. Studia i prace wydziału nauk ekonomicznych i zarządzania. N 15, 2018, p. 157–167.
5. Kowgier H. Kilka uwag o wymiarze fraktalnym Minkowskiego oraz wykładniku Hursta na giełdzie papierów wartościowych. Studia i prace wydziału nauk ekonomicznych i zarządzania NR 15, 2016, str. 157–167.
6. Orzeszko W. Wymiar fraktalny szeregów czasowych a ryzyko inwestowania / Acta Universitatis Nicolai Copernici Ekonomia Xli – Nauki Humanistyczno-Społeczne – Zeszyt 397, Toruń 2010, P. 57–70.
7. Milton S. Raimundo. Application of Hurst Exponent (H) and the R/S Analysis in the Classification of FOREX Securities. International Journal of Modeling and Optimization 8(2): 116-124 DOI: 10.7763/IJMO.2018.V8.635
8. Feder E. Fraktaly / Feder E. ; per. s angl. – М. : Mir, 1991. – 254 с.
9. Zwolankowska M. Metoda segmentowo-wariacyjna. Nowa propozycja liczenia wymiaru fraktalnego // Przegląd Statystyczny, 2000, R. 47, z. 1-2, S. 209–224.
10. Skrzat J. Accuracy of the box-counting algorithm for noisy fractals. International Journal of Modern Physics C December 2014.

DOI: 10.1142/S0129183116501126 – <https://www.researchgate.net/publication/269935552>

11. Lukach M.O. Typy linhvistychnykh modelei ta yikh zastosuvannia dlia rozviazannia linhvistychnykh zadach, 2016. – URL : http://science.lp.edu.ua/sites/default/files/Papers/19_151.pdf
12. Vasyliiev O. Matematychni metody y modeli v linhvistytsi / O. Vasyliiev, O. Chalyi, I. Vasyliieva. – URL : https://uamoderna.com/images/archiv/27-2020/27_9_28%20Oleksii%20VASYLIEV_compressed.pdf
13. Vavilenkova A. Pobudova zmistovnoi modeli tekstu na osnovi vykorystannia lohiko-linhvistychnykh modelei. – Lviv Polytechnic National University, 2016. – S. 169–175. – URL : <http://ena.lp.edu.ua:8080/bitstream/ntb/31313/1/25-169-175.pdf>
14. Kalymon Yu.O. Modeliuvannia: linhvistychnyi aspekt / Yu.O. Kalymon // Molodyi vchenyi. – 2019. – № 4.2 (68.2). – С. 88–91.
15. Bol'shakov I.A. Prikladnaya teoriya sluchajnykh potokov / Bol'shakov I.A., Rakoshic V.S. – M. : Sov. Radio, 1978. – 248 s.
16. Landeh D.V. Fraktal'nye svoystva tematicheskikh informacionnykh potokov iz Internet / D.V. Landeh // Reestraciya, zberigannya i obrobka danikh. – 2006. – T. 8, № 2. – S. 93–99.
17. Ivanov S.A. Statisticheskij analiz dokumental'nykh informacionnykh potokov / S.A. Ivanov, N.V. Krukovskaya // NTI. SER. 2. INFORM. PROCESSY I SISTEMY. – 2004. – № 2. – S. 11–14.
18. Landeh D.V. Osnovy integracii informacionnykh potokov : monografiya / Landeh D.V. – K. : Inzhiniring, 2006. – 240 s.
19. Bohdan S. K. Metody y metodyka linhvostylistychnykh doslidzhen: metodychni rekomendatsii dlia slukhachiv i kerivnykiv seksii ukrainskoi movy / S. K. Bohdan. – Lutsk, 2011. – 28 s.
20. Kulchytskyi I.M. Doslidzhennia dovezhny rechennia ta slova u tvorakh Romana Ivanychuka. – URL : <http://science.lpnu.ua/sites/default/files/journal-paper/2018/jun/13011/ilovepdfcom-139-148.pdf>
21. Zaveriushchenko M.P. Ofitsiino-dilovyi styl: pravyla ukladannia dokumentiv riznykh vydiv : navch. posib. shchodo samostiinoi roboty studentiv / M.P. Zaveriushchenko, O.M. Krymets, S.M. Cherniavska, O.V. Shokurov. – Kharkiv : NTU «KhPI», 2019. – S. 93.

Надійшла/Paper received : 04.03.2021 р. Надрукована/Printed : 02.06.2021 р.