https://doi.org/10.31649/1997-9266-2023-167-2-58-65
УДК 519.722:621.391

В. Г. Крижановський[1]

ЕНТРОПІЯ ТА КІЛЬКІСТЬ ІНФОРМАЦІЇ У ТЕХНІЧНИХ ПОЗНАЧЕННЯХ

[1]Донецький національний університет імені Василя Стуса, Вінниця

Розглянуто умовні позначення інтегральних мікросхем як приклад класифікації та скороченої назви (коду) технічних виробів для відповіді на запитання: Чому кажуть, що деякі системи позначень є «інформативнцішими?». Чи дійсно в таких позначеннях міститься більше інформації порівняно з іншими системами? Такі задачі тісно пов'язані з задачами машинного навчання та побудови «семантичної павутини». На основі алгебраїчного підходу та теорії множин розглянуто характеристики ентропії класифікації позначень та показано, що ентропія такого кодованого позначення менше ніж довільної системи запису технічних характеристик, що пояснюється позиційною структурою позначення і відповідно меншою потужністю множин, які складають конкретне позначення. На основі підходу інформаційної алгебри підтверджено, що встановлення в технічних позначеннях атомарної структури множин, яким ставляться у відповідність технічні характеристики, дійсно відповідає математичному визначенню інформативнішої структури. На основі математичної теорії натяків проаналізовано структуру технічного позначення та вказано на можливість отримання додаткової інформації, наприклад взаємозв'язків між різними групами технічних параметрів. Вона буде отримана внаслідок запитань, що уточнюють інтерпретацію існуючих відповідей. Це є наслідком властивості ентропії натяків, яка має дві складові — ентропію Шеннона та узагальнену міру Хартлі, які відповідають ймовірнісній інформації про справжню інтерпретацію відповіді в наборі та реляційній інформації про справжню відповідь про деякий тип параметрів інтегральних схем. Технічне позначення виявляється дієвим прикладом, на якому можна застосувати розглянуті математичні теорії, і відповідно може бути прикладом коду, який, з одного боку може бути зрозумілим людині, а з іншого — використовуватися в системах машинної обробки інформації.

Ключові слова: ентропія, інформація, технічні позначення, алгебраїчна теорія ентропії, ентропія класифікації, інформаційна алгебра, теорія натяків (підказок).

Вступ

Умовні позначення виробів, деталей та обладнання у техніці є поширеною практикою. Ці позначення часто є системою класифікації об'єктів, та такими «що говорять», тобто у самих позначеннях окрім ідентифікації виробів є і характеристика їхніх основних технічних параметрів. Виникає запитання — а яка саме кількість інформації міститься у цих умовних позначках та чи дійсно у позначеннях «що говорять» закладено більше інформації? Чи є в системі таких позначень перевага в кількості інформації (або в кількості інформації на один символ позначення) порівняно з порядковим номером у каталозі виробів, якщо все одно в більшості випадків потрібно звертатися до технічного опису виробу?

Ця задача тісно пов'язана з задачами побудови «Семантичної павутини» (semantic Web) [1], машинного розпізнавання тексту (побудови баз знань) [2], класифікації понять та багатьох інших, в яких потрібно обробляти інформацію, що може бути віднесена до поняття «знань» [3]. У цих ситуаціях проводиться розробка коду, який в стислому вигляді передає якусь інформацію про технічний об'єкт. Далі цей код використовується як для обробки так і передачі інформації. Кожному об'єкту відповідає одне слово коду, яке є окремим повідомленням.

Є різні умовні позначення, які одночасно є системами класифікації, до них належать міжнародна патентна класифікація (МПК), універсальна десяткова класифікація документів (УДК), коди ідентифікації книжок (ISBN), журналів (ISSN) або авторських прав (DOI). Є коди, які оптимізовані для машинної обробки — штрих-коди, QR-коди, і є такі, які одночасно можуть розуміти і люди — мова Асемблер, МПК та інші. Однак не для всіх цих кодів можна встановили однозначну відповідність (наприклад МПК та УДК є багатозначні), DOI за визначенням має строгу відповідність.

Серед цих кодів (систем позначень) можна виділити коди, у яких є властивість «говорити», тобто надавати деяку частину потрібної інформації без процедури декодування — звернення до таблиць відповідності і технічних описів виробу. Розглянемо їхню об'єктивну характеристику — ентропію. Як приклад виберемо систему кодування виробів електронної техніки — інтегральних мікросхем (ІС) [4]. Ця система передбачає умовне позначення ІС у вигляді літерно-цифрового коду, складеного за певними правилами, які хоча і змінювались, зберігають основні принципи побудови і продовжують і зараз використовуватися в Україні. Ця система позначень ІС є моделлю для розгляду аналогічних систем в різних галузях науки і техніки.

Оскільки оцінити кількість інформації в такому випадку буде вкрай важко внаслідок того, що інформація буде залежати від отримувача (його апріорної інформації), то будемо аналізувати саме ентропійні характеристики позначення. Деякі аспекти залежності кількості інформації від отримувача розглядаються в [5]. Головним об'єктом дослідження буде зміна ентропії внаслідок класифікації і структурування характеристик в позначенні, тому що в цьому сенсі різниця ентропії дорівнює кількості інформації, яку можна отримати [6].

Метою роботи є аналіз інформаційної ємності систем позначень об'єктів з певної множини на основі ентропійного підходу та виявлення їхніх властивостей з метою побудови систем виявлення інформації про характеристики цих об'єктів та зв'язків між ними.

Приклад технічного позначення

На рис. 1 показано приклад умовного технічного позначення (ТП) інтегральної схеми (ІС), Воно має стандартизований вигляд (стандарт 1980 р.) и складається з п'яти груп зі змінною кількістю букв (Б) або цифр (Ц): **1** — $(0\vee 1)$Б + $(1\vee 2)$Б; **2** — 4Ц; **3** — 2Б; **4** —2Ц; **5** — 1Б, тобто у групах може бути різна кількість символів, також можуть бути додані додаткові символи в кінці позначення [4]. Без втрати загальності вважаємо, що у першій групі дві букви. Таким чином, маємо п'ять груп, з яких три літерні, а дві цифрові. Оскільки для спрощення будемо розглядати позначення поодинці, то можна припусти що всі символи однаково ймовірні, але для літерних позначень введемо приблизну кількість можливих комбінацій. Згідно зі стандартом 1980 року у першій групі символ 1b може бути або 2 або К, у той час як 1c може набувати значення з 9 літер [4]. Елемент 2a вибирається з 8 цифр, а 2b — з 9 і 2c може набувати значення від 01 до 99. Елементи 3 можуть набувати 140 значень, тоді як 3a може набувати 19 значень, а 3b — 16, 4 будемо вважати від 01 до 99), елемент 5 можна вважати таким, що вибирається з 30 літер. Оскільки процес класифікації і ТП ІС постійно розвивається то всі цифри даються приблизно, для ілюстрації прикладів. У табл. 1 показано ТП, що входять у код ІС, що об'єднано у 8 підгруп, за якими буде проводитися аналіз.

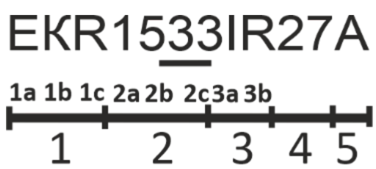

Рис. 1. Структура умовного позначення ІС

Таблиця 1

№	Позначення	Призначення підгруп	Кількість символів
1	1b	Виконання	2
2	1c	Тип корпусу	9
3	2a	Тип технології	8
4	2b	Серія мікросхем 1	9
5	2c	Серія мікросхем 2	99
6	3a	Тип мікросхеми	140
7	3b		
8	4	Номер розробки	99
9	5	Інші характеристики	30

Умовні позначення є кодом, в якому є взаємно-однозначна відповідність (бієкція) між відповідною технічною характеристикою та елементом в умовному позначенні. Для використання алгебраїчного підходу до обчислення інформаційних характеристик позначення введемо універсальну множину *U*, підмножиною якої будуть всі технічні характеристики умовного позначення. І відповідно елементи коду, які відображають ці характеристики, також будуть підмножинами універсальної множини. Введення універсальної множини дозволяє використовувати теорію множин для розрахунку інформаційних параметрів умовних скорочень. Множина позначень утворює класифікацію всієї кількості об'єктів (ІС) за їхніми технічними характеристиками.

Ентропія позначення

«Інформаційний зміст» класифікації розглянуто в роботі [7], де на основі алгебраїчного підходу розраховується ентропія як масиву класифікованих величин, так і одного варіанту класифікації. Оскільки для розуміння результатів будуть потрібні особливості розгляду, то надаємо в скороченому вигляді деталі виведення результату.

Класифікація, з позиції теорії множин, для деякої множини X — це множина підмножин $\{B_1,....,B_n\}$ з X, які загалом не повинні бути непересічними [7]. Зрозуміло, що приклади системних позначень, які визначені вище, підпадають під це визначення. Але потрібно вказати правило складання елементів B_i та відношення між ними, щоб обчислити ентропію. Тут X можна трактувати як набір сукупностей технічних характеристик всіх існуючих ІС, які мають умовні позначення. Введемо деякі поняття, важливі для подальшого розгляду.

Міра — міра m на скінченній множині А є функцією $m : \mathcal{P}(A) \to \mathbb{R}^+$ такою, що для будь-яких двох непересічних елементів S, $T \in \mathcal{P}(A)$ виконується $m(S \cup T) = m(S) + m(T)$. Тут $\mathcal{P}(A)$ означає множину всіх підмножин цієї множини A (булеан). Кількість елементів у множині А з n елементів (потужність) — $|A| = n$. Норма множини (колекції) C з підмножин $\{B_1,....,B_n\}$, які складаються з елементів X, $C = \{B_1,...,B_n\}$ визначається $\|C\| = \sum_{i=1}^{n} |B_i|$. Міри на наборах підмножин множини — m: $\mathcal{P}((\mathcal{P}(X)) \to \mathbb{R}^+$, також визначається нормалізована міра як часткове відображення μ: $\mathcal{P}((\mathcal{P}(X)) \to \mathbb{R}^+$ і записується як $\mu(C) = m(C)/\|C\|$, якщо $\|C\| \neq 0$.

Нехай $\alpha : X \to Y$ — відображення, де $X, Y \in U$. Якщо C — колекція, $C \subseteq \mathcal{P}(X)$, колекція $\alpha(C) \in \mathcal{P}((\mathcal{P}(X))$ визначається як $\alpha(C) = \{\alpha(B) \mid B \in C\}$. Якщо є сімейство мір M, таке, що $m_X(C) \geq m_Y(\alpha(C))$ для будь яких $X, Y \in U$ і будь яких відображень $\alpha : X \to Y$ та колекцій $C \in \mathcal{P}(X)$, то воно зветься таким, що скорочує.

В роботі [7] визначається сімейство ентропійних мір, що скорочує, на множині U як $\{m_X : \mathcal{P}(\mathcal{P}(X)) \to \mathbb{R}^+ \mid X \in U\}$, якщо $C \in (\mathcal{P}(X))$ і $D \in (\mathcal{P}(X))$, то $m_{X\times Y}(C\times D) = m_X(C) + m_Y(D)$ для всіх колекцій C, D таких що $\|C\| > 0$ і $\|D\| > 0$.

Визначальна умова для ентропійних сімей мір виражена в термінах нормалізованих мір як

$$\mu_{X\times Y}(C \times D) = \mu_X(C) + \mu_Y(D)$$

для всіх наборів C, Z, таких, що $\|C\| > 0$ і $\|D\| > 0$. Тут $\mu : (\mathcal{P}(X)) \to \mathbb{R}^+$ дається як, $\mu[C] = m(C)/\|C\|$, якщо $\|C\| \neq 0$.

Також в роботі [7] доведена теорема, яка для $k = 1$ вводить ентропію колекції.

Нехай $\{m_x : \mathcal{P}(\mathcal{P}(X)) \to \mathbb{R}^+ \mid X \in U\}$ — ентропійна сім'я мір на U. Для кожного набору $C \in \mathcal{P}(\mathcal{P}(X))$, такого, що $C = \{B_i | 1 \leq i \leq n\}$ маємо $m_X(C) = k \sum_{1\leq i\leq n} |B_i| \log_2 |B_i|$ для деякої невід'ємної дійсної сталої k.

Наведені вище результати легко узагальнити на випадок мультимножин. Мультимножина множини A є підмножиною M множини $A \times N$ таких, що для кожного $a \in A$ коли $n \in N$ так що $(a, m) \in M$ тоді і тільки тоді, якщо $m < n$. Множина всіх мультимножин A позначається $\mathcal{M}(A)$ [8]. У [7] визначено мультинабір X як мультипідмножину $\mathcal{P}(X)$. Для множини X ентропія мультиколекції $C = \{B_1,...,B_n\}$ підмножин X задається формулою 3 з [7]

$$H_X(C) = \sum_{1\leq i\leq n} |B_i| \log_2 |B_i|. \tag{1}$$

Числові характеристики ентропії для позначень ІС

Максимальна кількість ІС, які можна класифікувати за допомогою системи класифікації на рис. 1 з параметрами в табл. 1, становить $l = 2 \cdot 9 \cdot 8 \cdot 9 \cdot 99 \cdot 140 \cdot 99 \cdot 30 = 5{,}335 \cdot 10^{10}$. Якщо зробити оцінку за формулою з твердження 8 з [7] де $|X| = l = 5{,}335 \cdot 10^{10}$, а $\|C\| = r = 396$, то ентропія такої колекції

буде
$$H_X(C) \le r\log_2 l = 1{,}411 \cdot 10^4. \qquad (2)$$

Якщо застосована конкретна класифікація, то зменшення ентропії буде визначатися формулою (2) з [7] $\Delta H = \sum_{i \in C} |B_i| \left(\log_2|X| - \log_2|B_i|\right)$. В нашому випадку

$$\Delta H = \sum_{i-1}^{8} |B_i| \left(\log_2 l - \log_2|B_i|\right) = 1{,}157 \cdot 10^4. \qquad (3)$$

Відповідно ентропія мультиколекції $C = \{B_1,...,B_n\}$ для позначення ІС є

$$H_X = \sum_{i=1}^{8} |B_i| \log_2|B_i| = 2{,}541 \cdot 10^3. \qquad (4)$$

Цей результат пояснюється тим, що у визначенні цієї колекції зафіксовано лише кількість властивостей (підгруп) класифікації та їхня кількість у кожній групі, але розглядаються довільні об'єкти з множини *X*. Якщо в нас простори кожної групи не перетинаються, є незалежними, тоді ентропію за умови рівноймовірності можна записати як міру (ентропію) Хартлі

$$H_H = \sum_{n \in A} \log_2 D_n = 35{,}635. \qquad (5)$$

Якщо ймовірності будуть відрізнятися, тоді $H(D) \le \log_2|D|$. Видно велику різницю в ентропії одного запису (5) від ентропії колекції за незалежних наборів параметрів, коли вони входять тільки один раз і за наборів, в яких параметри можуть входити в класифікацію в довільних кількостях. Тому вже цей факт може свідчити про «інформативність» ТП, оскільки за такої класифікації ентропія визначення технічних характеристик істотно зменшується. Звісно це відбувається за рахунок апріорної інформації про позиції технічних параметрів у записі.

В інформаційній алгебрі кажуть, що інформація ϕ є інформативнішою, ніж ψ, якщо остання, поєднана з першою, не змінює її [9]. В ТП, що розглядається, внаслідок структуризації інформація, отримана з іншої групи, не змінює значення вибраної групи, тому окремі групи позначення є найінформативнішими. Друге підтвердження — це наявність атомарних пропозицій або атомів, це дрібніші запитання ніж інші, і відповідно до інформаційної алгебри вони ставлять точніше запитання і є інформативнішими. Найінформативнішими є атомарні судження, які мають бінарну відповідь «так» чи «ні». Це означає, що жодна інформація в домені, крім нульової інформації, не може бути інформативнішою ніж атом; ніяка інша інформація, крім нульової інформації, не може міститися в атомі. Це просто пояснити з погляду ентропії — апостеріорне значення ентропії при отримані інформації з атому є нульовим. Тому інформаційна алгебра підтверджує більшу інформативність ТП саме в сенсі цієї теорії [9]. З іншого боку, можна розглянути «питому ентропію», яку можна вивести за умови, що ми обчислюємо ентропії кількох множин, що не перетинаються (5), як $H_{HN} = \sum_{n \in A} \left(\log_2 D_n / n\right)$. Такі залежності зустрічаються у характеристиках обмежених послідовностей [10] та інформаційної ємності систем числення [11]. Графік залежності $\log_2 n/n$ показано на рис. 2.

Застосування теорії натяків до технічних позначень

Теорія натяків зараз розвивається як розгалуження теорії доказів Демпстера–Шейфера [12], і від неї очікують відображення простору ймовірності в звичайну систему розпізнавання [13]. Розглянемо на прикладі, які можливості відкриваються у застосуванні теорії натяків при розгляді нашого прикладу позначення. Важливим методичним висновком теорії натяків є формалізація множини можливих інтерпретацій, яка обов'язково присутня у механізмі логічних виводів, тобто у процесі набуття інформації. І відповідно наявність такої множини призводить по появи відповідної складової ентропії, яка не завжди розглядається у разі обчислення кількості інформації [13].

Натяк є частиною невизначеної інформації. Кожен натяк відноситься до *запитання* зі скінченним набором можливих *відповідей*, які називаються *фреймом розпізнавання*, у нас це сенс, який притаманний окремому прикладу позначення (рис. 3). Припускається, що тільки один з елементів множини Θ є істиною, але невідома відповідь. Другий компонент натяку — це скінченний набір Ω можливих *інтерпретацій*. Кожна інтерпретація обмежує можливі відповіді в межах Θ. Якщо $\omega \in \Omega$ є правильною інтерпретацією, тоді правильна відповідь має належати до деякої підмножини $\Gamma(\omega) \subseteq \Theta$, де $\Gamma: \Omega \to \mathcal{P}(\Theta)$ є багатозначним відображенням інтерпретацій Ω на набір потужностей $\mathcal{P}(\Theta)$ фрейму розпізнавання. Множина Γ(ω) називається фокальною множиною інтерпретації $\omega \in \Omega$. Однак не всі тлумачення однаково ймовірні. Тому припускається розподіл ймовірностей p, який призначає ймовірність $p(\omega) \geq 0$ кожній інтерпретації $\omega \in \Omega$. Загалом натяк $\mathcal{N}$ відповідає четвірці $\mathcal{N} = (\Theta, \Omega, p, \Gamma)$ з (Ω, p) як імовірнісного простору [13].

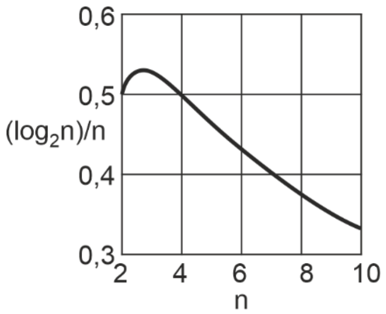

Рис. 2. Функція, яка дозволяє обчислити питому ентропію множини

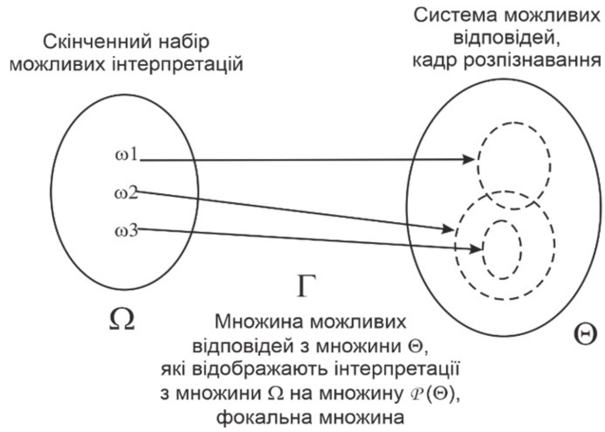

Рис. 3. Модель підказок, які виникають під час аналізу умовного позначення ІС, за [10]

Використовуючи результати [13], для інформації, яка міститься у підказці, напишемо вираз

$$i(\mathcal{N}) = \left(\log|\Omega| - H_{\mathcal{N}_N}(\Omega)\right) + \left(\log|\Theta| - GH(\mathcal{N}_N)\right), \qquad (6)$$

де $\mathcal{N}_N$ — нормалізований натяк, для якого $\sum\limits_{\omega\in\Omega:\Gamma(\omega)\neq\varnothing} p(\omega) = 1$;

$$H(\mathcal{N}_N) = -\sum_{\omega\in\Omega} p_N(\omega)\log_2(p_N(\omega)) \text{ —} \qquad (7)$$

— ентропія Шеннона для інтерпретацій,

$$GH(\mathcal{N}_N) = \sum_{A\subseteq\Theta} m(A)\log_2|A| \text{ —} \qquad (8)$$

— міра Хартлі для множини A, для ненульових натяків $|A| = |\Gamma(\omega)|$.

Ця ентропія відповідає умовній ентропії Шеннона, коли ми можемо отримати інформацію про елементи θ множини Θ, з елементів $\omega \in \Omega$. Це наслідок правила ланцюга для ентропії Шеннона [6]. Постулюється, що натяк містить два типи інформації: існує *ймовірнісна інформація* про справжню інтерпретацію в наборі та *реляційна інформація* про справжній елемент у кадрі розрізнення під кожною інтерпретацією [13].

Розрахуємо ентропію натяку, який може бути перетворений в справжнє висловлювання і таким чином зменшити ентропію ТП. Розглянемо ТП з рис. 1 — КR1533IR27А. Які натяки може побачити тут умовний фахівець, що не має під рукою довідника, але має деякий досвід у роботі з ІС. По-перше, буква «К» в першій позиції, фактично це не натяк, а визначення, тому що інші значення в цій позиції зустрічаються рідше, а в наших припущеннях існує тільки два варіанти, або є цей символ, або його немає. Тому ентропія цього елементу формально у розрахунку (5) дорівнює 1 біту, з формули (4) ентропія також дорівнює 1, і праві дужки у (6) дорівнюють 0. Тоді інформація цього натяку буде

$$i(\mathcal{N}) = \left(\log|\Omega| - H_{\mathcal{N}_N}(\Omega)\right) + \left(\log|\Theta| - GH(\mathcal{N}_N)\right) = \log_2 2 - \left(-\sum_{i=1}^{2} p_i \log_2 p_i\right).$$

І за ймовірності появи символу «К», близькому до 1, інформація натяку буде близька до 1 біту. І це зрозуміло, оскільки прийнято, що у множині Θ міститься правильна відповідь. А от якщо допусти однакову ймовірність появи та відсутності символу «К», то інформація натяку буде дорівнюватися нулю.

Для другої групи, літера «R» позначає тип корпусу ІС, — це пластмасовий корпус з дворядним розташуванням виводів. В нас ця підгрупа має 9 градацій, і обчислюючи інформацію натяку за формулою (6), отримаємо також рівність нулю другої пари дужок, а інформація підказки буде відповідати уточненню ентропії Шеннона відносно ентропії Хартлі, яку використано у (5).

Де натяк буде відрізнятися, так це для третьої групи, — тип ІС. Тут можна сформулювати різні запитання, що будуть або схожі на розглянуті для попередніх груп, або такі, що визначають ентропію підказки, яка порівнюється з величиною $\log_2 |\Omega \times \Theta|$. Наприклад, запитання: «Які типи ІС є в серії 1533?». Щоб отримати цю інформацію, потрібно скористатися апріорною інформацією про систему позначень типів ІС. Якщо підрахувати кількість різних типів за стандартом [4], то можна вважати $|\Omega| = |\Theta| = 140$. Але якщо розглядати збільшений поділ, за першою літерою, то $|\Omega| = |\Theta| = 19$. В серії 1533 налічується 5 різних перших літер в типах ІС [14]. Зауважимо, що запитання стосується числа сполучень із 19 по 5 елементів, це також пояснює збільшення ентропії у розгляді натяків. В такому варіанті ентропію Шеннона потрібно розглядати на просторі 19 елементів, а міру Хартлі на просторі 5, оскільки в нас тільки 5 можливих відповідей. Тоді, маючи апріорну інформацію, що серія відноситься до цифрових (логічних, обчислювальних) ІС, можна вибрати гіпотетичний розподіл ймовірностей з більшими значеннями для цих груп, і за формулою (3) отримати значення ентропії від 4,248 — для однакової ймовірності всіх 19 літер до 2,45 за реалістичнішого розподілу у припущені певних класів ІС. Якщо прийняти гіпотезу про рівномірний розподіл ймовірностей у множині Ω, тоді ліві дужки будуть прирівнюватися нулю, і залишиться лише права частина, яка буде показувати зменшення ентропії (інформацію) про вибірку лише 5 конкретних елементів з 19-ти. Ентропія Хартлі

$$GH\left(\mathcal{N}_N\right) = \sum_{A \subseteq \Theta} m(A) \log_2 |A| = 2{,}322 .$$

І інформація натяку відносно запитання «Які типи ІС є в серії 1533?» буде $i\left(\mathcal{N}\right) = 3{,}723$. Тут треба врахувати, що ми розглядаємо появу різних 5 символів, і тоді ентропія Хартлі буде 20,412 і це значно більше інформації у натяку, тобто натяк дає нам якусь частину інформації, але не всю. Щоб отримати повну інформацію у відповідь на запитання які саме типи ІС містяться у серії 1533, треба задавати дрібніші запитання стосовно кожної літери в позиціях 3a і 3b.

Практична цінність натяку полягає в тому, що він зменшує невизначеність позначення і збільшує ймовірність отримати правильну інформацію з нього. І ще, фактично натяк додає інформацію, яка виходить за рамки просто класифікації, а дозволяє розраховувати інформацію з розширеного розгляду про взаємні відношення об'єктів. Умовні позначення, які виконують функцію класифікації, дають значно звужені множини Ω та Θ, і спрощують застосування теорії натяків до отримання знань з даних.

З використанням властивостей технічних позначень розглянемо ще дві системи — DOI, цифрового ідентифікатора об'єктів інтелектуальної власності, та позначення, які мають в основі ІС 74 серії [15], [16]. Розглянемо DOI — 10.1109/MMM.2022.3226632, тут не вказана перша частина ідентифікатора https://doi.org, яка є адресою сайту та бази даних International DOI Foundation. У другій частині (до косої риски) також стоїть незмінний показник — 10, який вказує на те, що це ідентифікатор, далі йдуть 4 або 5 цифр, які вказують на видавця, в третій частині стоїть довільне але унікальне і для видавця і для DOI позначення на конкретний об'єкт. В прикладі число 1109 це IEEE, а MMM — IEEE Microwave Magazine, рік видання та номер статті. Тут є надлишковість і за рахунок цього за мінімальної апріорної інформації можна отримати відомості про рік видання та журнал. Це позначення є таким, «що говорить», його питома ентропія може бути підвищена у разі прийняття уніфікованої системи нумерації статей в журналах.

Позначення ІС 74 (7400, або 54/74) ведуть історію з перших ІС TTL, і підтримуються кількома фірмами. Наприклад, SN74ALS245 — двонаправлений восьмибітний буфер на основі TTL логіки з діодами Шоттки з низьким енергоспоживанням виробництва Texas Instruments. На жаль, еволюційний характер розвитку системи ТП 74 серії привів до того, що кількість символів у кожній групі позначення може бути різною, до того ж та сама група може позначати різні характеристики в залежності від символів, що входять в неї (можуть характеризувати технології, напруги живлення та швидкодію), тому обчислювати ентропію стає складніше. Часто ентропію треба обчислювати для ТП взагалі. Розглянуті приклади підтверджують, що для підвищення інформативності ТП потрібно розбивати їх на чітко визначені групи з меншою кількістю символів.

Висновки

Показано, що технічне позначення (на прикладі позначень ІС) має значно меншу апріорну ентропію ніж довільна колекція технічних характеристик. Це для позначення фактично коду є наслідком структури, в якій систематизовано технічні характеристики виробу. Застосування алгебраїчного підходу до теорії інформації дає змогу обчислити це зменшення ентропії, тобто зменшення невизначеності. Інформаційна алгебра розкриває ще один механізм збільшення інформативності технічних позначень — це атомарна структура позначення, коли інформація отримується з дрібніших розділів переліку технічних характеристик. Застосування теорії натяків до технічних позначень на прикладі позначень інтегральних схем показує, що отримання узагальненішої інформації з технічних позначень практичніше, бо зменшуються множини інтерпретацій та фреймів розпізнавання. Через це розроблення комп'ютерних програм для обробки даних, представлених у вигляді позначень на основі систем класифікації, стає актуальнішим.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

[1] P. Hitzler, “A review of the semantic web field,” *Communications of the ACM,* no. 64 (2), pp. 76-83, 2021. https://doi.org/10.1145/3397512 .

[2] А. І. Катаєва, «Застосування баз знань до неструктурованої текстової інформації,» *Матеріали наукової конференції професорсько-викладацького складу, наукових працівників і здобувачів наукового ступеня за підсумками науково-дослідної роботи за період 2019–2020 рр.* Вінниця: ДонНУ, квітень–травень 2021 р., с. 324-326.

[3] А. Ю. Берко, О. М. Верес, і В. В. Пасічник, *Системи баз даних та знань.* Кн. 1, *Організація баз даних та знань,* навч. пос. Львів, Україна: НУ «Львівська політехніка», 2013, 680 с.

[4] ОСТ 11 073.915-*80, Мікросхеми інтегральні. Класифікація і система умовних позначень.* Чинний від 1 січня 1980 р.

[5] U. Eco, *The Open Work***,** transl. by Anna Cancogni: with an introduction by David Robey. Harvard University Press Cambridge, Massachusetts, 1989, 290 p.

[6] R. M. Gray, *Entropy and Information Theory*. Springer New York, NY, 2013, 355 p.

[7] K. Baclawski, and D. A Simovici, “A characterization of the information content of a classification,” *Information Processing Letters*, vol. 57, issue 4, pp. 211-214, 26 February 1996.

[8] P. Fejer, and D. Simovici, *Mathematical Foundations of Computer Science,* Springer, New York, 1990.

[9] J. Kohlas, “Information Algebras: generic structures for inference,” *Discrete Mathematics and Theoretical Computer Science,* Series ISSN 1439-9911. ISBN 978-1-85233-689-9 .

[10] A. Janssen, and K. Immink, “An Entropy Theorem for Computing the Capacity of Weakly – Constrained Sequences,” *IEEE Tran. on Information Theory*, vol. 46, no. 3, pp. 1034-1038, May 2000.

[11] Р. Н. Квєтний, П. П. Повідайко, М. М. Компанець, В. В. Гармаш, і Я. А. Кулик*, Арифметичні основи проектування мікропроцесорних систем*, навч. посіб. Вінниця: ВНТУ, 2017, 111 с.

[12] J. Kohlas, “*The mathematical theory of evidence — A short introduction,*” in:, J. Doležal, J Fidler, Eds. System Modelling and Optimization. IFIP, Springer, Boston, MA, 1996, pp. 37-53. https://doi.org/10.1007/978-0-387-34897-1 .

[13] M. Pouly, J. Kohlas, and P. Y. A. Ryan, “Generalized Information Theory for Hints,” *January 2013 International Journal of Approximate Reasoning,* no. 54(1), pp. 228-251. https://doi.org/10.1016/j.ijar.2012.08.004 .

[14] *ЦЕОМ. Інтегральні мікросхеми серії КР1533*. [Електронний ресурс]. Режим доступу: https://ksm.nau.edu.ua/arhitectura/files/ims1533.pdf. 21.01.23 .

[15] А. Д. Данілова, А. І. Радченко, і Т. М. Яцків, *Методичні рекомендації щодо впровадження цифрових ідентифікаторів у видавничий процес для періодичних видань Національної академії наук України*, ПА «Укрінформнаука», 3-е вид., перер. і доп. Київ: Академперіодика, 2019, 60 с.

[16] A. Saha, and N. Manna, *Digital Principles and Logic Design*. Infinity Science Press LLC, 2007, 505 p. ISBN: 978-1-934015-03-2.

Рекомендована кафедрою комп'ютерних систем управління ВНТУ

Стаття надійшла до редакції 7.03.2023

Крижановський Володимир Григорович — д-р техн. наук, професор, професор кафедри прикладної математики та кібербезпеки, e-mail: v.krizhanovski@donnu.edu.ua .

Донецький національний університет імені Василя Стуса, Вінниця

V. G. Kryzhanovskyi[1]

Entropy and Quantity of Information in Technical Designations

[1]Vasyl′ Stus Donetsk National University, Vinnytsia

Conventional designations of integrated microcircuits are considered as an example of classification and abbreviated name (code) of technical products to answer the question: Why do they say, that some designation systems are "more informative?". Do such notations contain more information compared to other systems? Such tasks are closely related to the tasks of machine learning and the construction of the "semantic web". Based on the algebraic approach and set theory, the characteristics of the entropy of the classification of designations are considered and it is shown that the entropy of such a coded designation is less than that of an arbitrary system of recording technical characteristics, which is explained by the positional structure of the designation and, accordingly, the lower power of the sets that make up a specific designation. Based on the approach of informational algebra, it is confirmed that the establishment in the technical notation of the atomic structure of the sets to which the technical characteristics correspond, really corresponds to the mathematical definition of a more informative structure. Based on the mathematical theory of hints, the structure of the technical designation is analyzed and the possibility of obtaining additional information, for example, relationships between different groups of technical parameters, is indicated. It will be obtained as a result of questions clarifying the interpretation of existing answers. This is a consequence of the property of hint entropy, which has two components — the Shannon entropy and the generalized Hartley measure, which correspond to probabilistic information about the true interpretation of the answer in the set and relational information about the true answer about some type of integrated circuit parameters. Technical notation turns out to be an effective example on which the considered mathematical theories can be applied and accordingly can be an example of a code that, on the one hand, can be understood by a person, and on the other hand, can be used in machine information processing systems.

Keywords: entropy, information, technical notation, algebraic theory of entropy, entropy of classification, information algebra, theory of hints.

Kryzhanovskyi Volodymyr G. — Dr. Sc. (Eng), Professor, Professor of the Chair of Applied Mathematics and Cybersecurity, e-mail: v.krizhanovski@donnu.edu.ua