

ІНФОРМАЦІЙНЕ СЕРЕДОВИЩЕ СИСТЕМ КОМП'ЮТЕРНОЇ ПІДГОТОВКИ ВИДАНЬ ТА ЙОГО ОСНОВНІ ХАРАКТЕРИСТИКИ

*Всеволод Сеньківський, Ігор Драбик, Роман
Ріпецький*

Вважається, що інформація — це основоположне поняття кібернетики, яке не потребує визначення, а розкривається на основі здійснюваних над нею перетворень. З часом змінювалися об'єми, суть, швидкість обміну, рівень сприйняття інформації. Її призначення набуває іншого змісту. Роль інформації в житті людей стає визначальною. Вона перетворюється в предмет всестороннього вивчення і дослідження, вимагаючи при цьому кардинально нових підходів та нетривіальних рішень, пов'язаних з її фіксацією, обробкою, передачею, зберіганням. Розв'язання кожної з цих задач обумовило необхідність постановки і вирішення складних проблем, таких, наприклад, як: оптимальне кодування, моделювання процесів перетворення інформації з використанням комп'ютерної техніки, визначення пропускну здатності каналів зв'язку та швидкостей передачі інформації, розробка спеціальних систем керування базами даних. Все це привело до появи нових наукових напрямів, принципово нових комп'ютерних інформаційних технологій, до зміни ролі і змісту діяльності людини в цих процесах.

Однією із основних характеристик інформації є її кількість. Своєчасне і точне визначення цієї величини має велике значення для нормального функціонування процесу, в якому використовується та чи інша інформація. Визначальними тут є асортимент, конфігурація та параметри технічних засобів, які генерують, фіксують, переробляють різні дані (сигнали), лінії зв'язку та передачі, запам'ятовуючі пристрої. Якщо процес автоматизований, то до них додаються вимоги до характеристик програм обробки інформації, об'ємів оперативної та зовнішньої пам'яті комп'ютерів, їх швидкодії і надійності.

В 1928 році американський інженер Р. Хартлі ввів двійкову логарифмічну міру для визначення кількості інформації:

$$I = \log_2 N,$$

де N - кількість станів інформації. Формула не є остаточною, тому що не враховує ймовірності появи того чи іншого стану. Кількість інформації залежить від числа можливих сигналів, повідомлень, символів (букв, знаків для тексту), яке вносить певну невизначеність. Ця апріорна невизначеність є тим більшою, чим більшим є число повідомлень. А це значить, що кількість інформації стає більшою, коли невизначеність знімається.

Крім кількісних показників, інформація характеризується також показниками якісними. Основними серед них є тип даних, що визначає смисловий зміст інформації. Тип даних визначається множиною всіх значень, які можуть приймати дані цього типу, і набором операцій, які можуть виконуватися над даними цього типу.

З розвитком інформаційних процесів вирішальною їх складовою стає *текстова інформація як поєднаний відповідним змістом текст, доповнений командами поліграфічного оформлення, певним способом закодований та записаний на комп'ютерному носії*. Текстова інформація має ті ж властивості, що і її першооснова — початковий текст: семантично-ієрархічну структуру, складність, що визначається відповідними групами, мовні особливості. Отже, текстова інформація — це не просто текст, а вхідні дані для комп'ютера, тобто компонент, який можна опрацювати і досліджувати згідно закономірностей, характерних для будь-якої інформації.

Ми свідомо звужуємо вхідні дані для систем комп'ютерної підготовки видань (СКПВ), обмежуючи їх тільки текстовою інформацією. В дійсності загальна вхідна інформація багатьох комп'ютерних видавничих систем включає графічні об'єкти для широкого спектру графічної та образотворчої продукції. Крім компонування таких елементів, сюди відносяться задачі растровання, кольороподілу, коректування і т. п. Все це становить важливу проблему, вирішення якої виходить за рамки даної роботи. Треба, однак, відзначити, що підходи та методи, запропоновані для текстових даних, можуть бути використані для інформації іншого типу, яка є вхідним елементом для систем комп'ютерної підготовки видань.

При підготовці текстової інформації використовується відповідне шрифтове забезпечення, яке містить букви, цифри, математичні символи, спеціальні знаки.

Певний набір знаків, який забезпечує випуск тих чи інших типів видань, комплектується у вигляді *шрифтової гарнітури*. Основний комплект знаків текстових гарнітур включає знаки

українського (раніше російського) та латинського алфавітів, цифри, спеціальні знаки. Шрифтове забезпечення складного набору додатково містить: знаки грецькі, готичні, індекси, знаки астрономічні та математичні, знаки структурних формул і т. п.

Множину споріднених між собою знаків шрифтової гарнітури назвемо алфавітом. З огляду на це текстова гарнітура складатиметься із чотирьох алфавітів: українського (U), латинського (L), цифрового (C) та алфавіту спеціальних знаків (Z).

Розв'язання задач кодування та обґрунтованого визначення інформаційних структур даних вимагає кількісних оцінок текстової інформації. Для їх визначення використовуємо ентропію, як одну з основних характеристик інформації:

$$H(T) = - \sum_{i=1}^N p_i \log_2 p_i. \quad (1)$$

Формула одержана за умови існування невизначеної ситуації T , яка представляє собою множину деяких станів t_1, \dots, t_N , ймовірності виникнення яких відповідно мають значення p_1, \dots, p_N , а невизначеності для кожного стану: $-\log_2 p_1, \dots, -\log_2 p_N$.

Для визначення ентропії сукупності символів текстової інформації застосовані загальні положення, прийняті в теорії інформаційних процесів. Відомо, що передача інформації не обмежується єдиним повідомленням. Найчастіше генерується послідовність, кожний елемент якої є підмножиною із множини можливих повідомлень.

Приймемо, що джерелом інформації в процесі комп'ютерної підготовки видання є СКПВ, яка генерує множину незалежних дискретних символів із загальної множини, що входить, наприклад, в алфавіт LG. Одночасно кожний символ є підмножиною алфавітів (U, L, C, Z) або послідовності сигналів, в яких визначені окремі значення

$$t_1^{(U)}, \dots, t_{N_U}^{(U)}, t_1^{(L)}, \dots, t_{N_L}^{(L)}, \dots, t_1^{(Z)}, \dots, t_{N_Z}^{(Z)}$$

з ймовірностями виникнення

$$p(t_1^{(U)}), \dots, p(t_{N_U}^{(U)}), p(t_1^{(L)}), \dots, p(t_{N_L}^{(L)}), \dots, p(t_1^{(Z)}), \dots, p(t_{N_Z}^{(Z)}).$$

Застосування формули (1) передбачає подати послідовність T_U, T_L, T_C, T_Z як деякий об'єднаний сигнал, число станів якого $K = N_U + N_L + N_C + N_Z$ із ймовірністю виникнення j -го стану сигналу ($1 \leq j < K$), рівного p_j .

Для спрощення перетворень використовуємо послідовність T_U, T_C

за складністю характерну для текстів художньої літератури. При цьому число станів $K_l = N_U N_C$. Позначивши через p_l ймовірність виникнення l -го стану сигналу ($1 < l < K_l$) одержимо ентропію алфавіту T_U, T_C :

$$H(T_U, T_C) = - \sum_l p_l \log_2 p_l.$$

Стан загального сигналу T_U, T_C характеризується значеннями $t_i^{(U)}, t_j^{(C)}$ для $i = 1, \dots, N_U; j = 1, \dots, N_C$.

На основі теореми добутку ймовірностей одержимо:

$$\begin{aligned} H(T_U, T_C) &= - \sum_{i=1}^{N_U} \sum_{j=1}^{N_C} p(t_i^{(U)}) p(t_j^{(C)}) \log_2 [p(t_i^{(U)}) p(t_j^{(C)})] = \\ &= - \sum_{i=1}^{N_U} p(t_i^{(U)}) \log_2 p(t_i^{(U)}) - \sum_{j=1}^{N_C} p(t_j^{(C)}) \log_2 p(t_j^{(C)}) = \\ &= H(T_U) + H(T_C). \end{aligned}$$

Аналогічно доведено, що для максимально складного видання ентропія текстової інформації рівна:

$$\begin{aligned} H(T_U, T_L, T_C, T_Z, T_M, T_G, T_I, T_S) &= H(T_U) + H(T_L) + \\ &+ H(T_C) + H(T_Z) + H(T_M) + H(T_G) + H(T_I) + H(T_S), \end{aligned} \quad (2)$$

де T_U, \dots, T_S — послідовності знаків алфавітів.

Розглядається система комп'ютерної підготовки видань, яка передає текстову інформацію T у фотовивідну секцію, котра в результаті дії перешкод в каналі приймає текст Z , неідентичний текстові T . У даному випадку використовується величина $p(t/z)$, яка визначає ймовірність передачі знаку t при умові, що фотовивідна секція прийме знак z . Тоді кількість інформації під час передачі одного знаку тексту

$$I(z_j, t_i) = H(t_i) - H(t_i / z_j) = \log[p(t_i / z_j) / p(t_i)]. \quad (3)$$

Середня кількість текстової інформації, одержаної при неповній достовірності передачі рівна:

$$I(Z, T) = H(T) - H(T / Z), \quad (4)$$

де $H(T/Z)$ — спільна ентропія вхідного і вихідного текстів, яка визначає ступінь втрати інформації в каналі і має вигляд:

$$H(T/Z) = - \sum_{i=1}^n \sum_{j=1}^n p(i_i, z_j) \log_2 p(i_i, z_j). \quad (5)$$

Враховуючи, що середній час проходження одного символу рівний τ_T , швидкість передачі інформації від комп'ютера до фотовивідної секції

$$V_T = H(T) / \tau_T = V_{T_{cp}} H(T), \quad (6)$$

де $V_{T_{cp}}$ — середня швидкість одержання символів від комп'ютера.

Максимальна швидкість одержання кодів фотовивідною секцією при фіксованій швидкості передачі в каналі без перешкод

$$C = V_{T_p} \max[H(T)]. \quad (7)$$

Для каналу з перешкодами одержимо відповідно

$$V(Z, T) = V_{T_{cp}} [H(T) - H(T/Z)] \quad (8)$$

та

$$C = V_{T_{cv}} \max[H(Z) - H(Z/T)] = V_{T_{cv}} \max I(Z/T). \quad (9)$$

Отже, пропускна здатність каналу з перешкодами визначається максимальною кількістю інформації, одержаної при неповній достовірності її передачі за фіксований проміжок часу.

Література

1. Бауэр Ф.Л., Гооз Г. Информатика. Вводный курс: В 2-х ч. Пер. с нем. - М.: Мир, 1990. Ч.1. - 336с., Ч.2. - 423с.
2. Верляль А.Ф., Широкин В.П. Информатика и ЭВМ. - К.: Техніка, 1987. - 344с.
3. Куликовский Л.Ф., Мотов В.В. Теоретические основы информационных процессов. - М.: Высш. шк., 1987. - 248с.
4. Основы теории информации и кодирования/ И.В.Кузьмин, В.А.Кедрус. - 2-е изд., перераб. и доп. - К.: Вища шк. Головное

изд-во, 1986. - 238с.

5. *Сеньковский В.Н.* Автоматизация фотонаборных процессов. - Львов: Вища шк., Изд-во при Львов. ун.-те, 1987. - 140с.

6. *Сеньківський В.М.* Автоматизація процесів формування тексту при комп'ютерній підготовці видань. Друга українська конференція з автоматичного керування "Автоматика-95". Тези доповідей. - Львів, 1995, ч.2, с. 65-66.