4.4 Інформаційні потоки

Інформаційним потоком від вузла S_i до вузла S_j інформаційної системи будемо називати функцію $i_{ij}(t)$, що задовольняє умови:

$$\int_{t}^{t+\tau} i_{ij}(t)dt = I_{ij}(\tau) \qquad 0 \le i_{ij}(t) \le v_{ij}, \tag{4.20}$$

де τ – час транзакції передачі інформації від вузла S_i до вузла S_j ; $I_{ij}(\tau)$ – кількість переданої інформації за час транзакції.

Якщо інформація передається в дискретному вигляді неподільними блоками (розмір блока визначається протоколом передачі: біт, байт, слово, пакет, повідомлення), то інформаційний потік обчислюється в середньому за період *транзакції*

$$i(t) = \frac{I_0(t,P)}{\Delta t}. \tag{4.21}$$

Якщо інформація передається в аналоговому вигляді, то інформаційний потік

$$i(t) = \frac{dI(t,P)}{dt}, \tag{4.22}$$

де I – кількість інформації; t – час її одержання; P – ймовірність працездатності системи.

Для моделювання інформаційних потоків у *документографічних системах* використовуються методи функціонально–операційного аналізу, аналізу норм вироблення рішення, модуль–метод, семіотичний аналіз, метод схем інформаційних зв'язків, матричний метод і ряд інших. Найбільше поширення одержала матрична модель.

Матрична модель будується у вигляді таблиці, що відбиває види документів, якими обмінюються підсистеми, перелік показників вхідних документів, що беруть участь у формуванні показників вихідних документів, частоту формування кожного документа, трудомісткість формування і значимість кожного показника та ряд інших допоміжних даних. Матрична модель перетворюється в орієнтований граф, у вершинах якого розташовані абоненти (процеси) інформаційної системи, а дуги відповідають переданим документам.

Аналіз інформаційних потоків зводиться до аналізу інтенсивності взаємодії виконуваних інформаційних процесів і сегментів інформації (документів). При цьому будують матрицю

$$Q_k^i = \begin{cases} 1, при\ B = істина \\ 0, при\ B = хибність, \end{cases} \quad (4.23)$$

де B=*істина*, якщо k–й сегмент використовується для реалізації i–го процесу.

На підставі матриці (4.20) будується матриця інтенсивностей використання інформаційних потоків

$$q_{ki} = n_k \cdot Q_k^i, \quad (4.24)$$

де n_k – частота використання сегмента.

В результаті побудови таких матриць і їх зіставлення з аналогічними матрицями вихідних інформаційних потоків мінімізують надлишковість інформаційних потоків.

Інформаційна модель *систем зв'язку* вперше була розроблена К. Шенноном і заклала основу теорії передавання інформації. Згодом теорія інформації була розвинута з метою поширення поняття інформації на інші галузі техніки. Відповідно до цієї теорії інформація передається у вигляді повідомлень від джерела до приймача каналом зв'язку за допомогою *інформативних сигналів*.

Фундаментальні співвідношення для оцінювання інформаційних характеристик *каналів зв'язку* послужили основою для створення інформаційних моделей інших видів інформаційних систем.

У теорії зв'язку розглядаються інформаційні аспекти передавання повідомлень дискретними каналами без завад і з завадами.

У *каналі з завадами* після прийому повідомлення залишається апостеріорна невизначеність стану $H_{X/Y}$ джерела інформації. Тому

$$I_{X,Y} = H_X - H_{X/Y} = -\sum_{i=1}^{n} p(z_i)\log_a p(z_i) + P_0 \log_a P_0 + (1-P_0)\log_a(1-P_0), \quad (4.25)$$

де P_0 – ймовірність правильного прийому повідомлення; n – число можливих станів X джерела повідомлень; $p(z_i)$ – ймовірність стану x_i; a – основа логарифму. При a=2 ентропія й інформація вимірюються в бітах.

Інформаційний потік у дискретному каналі зв'язку

$$i = F_0 \cdot I_{X,Y}, \quad (4.26)$$

де F_0 – гранична частота передавання елементарних повідомлень $I_{X,Y}$, яка залежна від фізичних характеристик каналу.

Канал зв'язку характеризується *пропускною спроможністю*:

$$C = F_0 \cdot \max[I_{X,Y}]. \quad (4.27)$$

Ймовірність правильного прийому P_0 залежить від характеристик завад у каналі зв'язку і способу кодування інформації. Існує безліч моделей каналів: гаусівські канали, канали з групуванням завад, канали із завмираннями тощо, що відрізняються статистичними характеристиками завад. Для кожного виду каналів використовуються методи кодування, що забезпечують максимум P_0.

Пропускна спроможність гаусівського каналу з завадами у вигляді білого шуму відповідає *теоремі К. Шеннона*

$$C = F_0 \cdot \log\left(1 + P_C / P_П\right), \tag{4.28}$$

де P_C – потужність інформативного сигналу; $P_П$ – потужність завад.

У моделях *систем телекомунікацій* інформаційні потоки задають на графі, вузли якого зіставляються з вузлами системи, а дуги – з каналами зв'язку. При цьому припускають виконання співвідношень:

1) вимога до *пропускної спроможності* каналу між вузлами k і l

$$\mathrm{i}_{\mathrm{kl}} \le C_{kl}; \tag{4.29}$$

2) вимога адитивності і передачі, що не спотворює інформацію

$$\sum_{i=1}^{n} i_{lk} = \sum_{r=n+1}^{n+m} i_{kr}, \tag{4.30}$$

де n – число потоків i_{lk} , що входять у вузол k; m – число потоків i_{kr} , що виходять з вузла k.

Наведені вимоги передбачають вимірювання інформаційних потоків у структурних одиницях – *словах, повідомленнях* або *пакетах* повідомлень рівної довжини, а також відсутність обробки інформації у вузлах системи. При використанні статистичної міри інформації вимога адитивності виконується тільки для незалежних повідомлень, що в реальних умовах експлуатації систем виконується не завжди.

Основні задачі, які розв'язуються в теорії телекомунікаційних систем, – це задачі *телетрафіку*. У цих задачах оцінюється інформаційне навантаження системи. Основи теорії телетрафіку закладені А. К. Ерлангом. Відповідно до цієї теорії телекомунікаційна система розглядається як система масового обслуговування. В результаті розв’язання задач телетрафіку визначають оптимальні пропускні спроможності каналів зв'язку, мінімальні маршрути при передаванні повідомлень між несуміжними вузлами, оцінюються характеристики ефективності обслуговування заявок (час проходження повідомлення в системі, ймовірність втрати повідомлення тощо).

Останнім часом питання інформаційних потоків глибоко розглядаються в рамках проектування і аналізу *протоколів передавання даних.*

Вимірювальні канали різноманітних автоматичних і автоматизованих систем призначені для одержання вимірювальної інформації. Поняття *вимірювальної інформації* було введене К.Б. Карандєєвим. Згодом зусиллями багатьох учених були створені основи інформаційно–енергетичної теорії вимірювань.

В інформаційно–енергетичній теорії вимірювань інформація розуміється в класичному статистичному розумінні як міра зменшення невизначеності апріорних знань про значення вимірюваної величини при одержанні результату вимірювання. В енергетичному аспекті розглядаються витрати енергії на одержання вимірювальної інформації.

Продуктивність джерела вимірювальної інформації (інформаційний потік)

$$i_{X,Y} = \frac{dI_{X,Y}(t)}{dt}. \qquad (4.31)$$

Інформаційні моделі *обчислювальних мереж* залежать від архітектурних особливостей їх побудови і протоколів передавання й обробки інформації. Архітектура обчислювальних систем і мереж зазнала значних зміни за період розвитку обчислювальної техніки. Найбільше поширення одержали локальні мережі зі спільним каналом, або моноканалом, через який абоненти мережі з'єднуються один з одним за принципом “кожен з кожним”. Абоненти обмінюються даними, розділеними на кадри. Формат кадру передбачає ряд полів. Усі поля містять кількість інформації, кратну 1 байтові. Байт – це найрозповсюдженіша в обчислювальній техніці одиниця кількості інформації в адитивній *структурній мірі* Хартлі. Відповідно до структурної теорії інформації основну роль у її вимірюванні відіграє запис інформації за допомогою комбінацій кодових символів – кодування. Найважливіші характеристики кодування – алфавіт, глибина і довжина коду. В обчислювальній техніці використовуються коди з глибиною 2 і алфавітом (0,1), або двійкові коди. Довжина коду визначається конструктивними особливостями елементної бази ЕОМ. Наразі при обміні даними між ЕОМ використовуються 8–розрядні коди. Це визначено стандартами на інтерфейси і протоколи обміну даними з метою забезпечення сумісності потоків даних з технічними засобами ЕОМ будь–яких типів. За допомогою 1 кодової комбінації з 8 розрядів можна закодувати $N=2^8$ даних. Відповідно кількість інформації в одній кодовій комбінації в структурній мірі

$$Q=\log_2 N = 8 \text{ бітів} = 1 \text{ байт}.$$

Структурна міра Q збігається зі *статистичною* I за умови рівної ймовірності кодових комбінацій і відсутності надлишковості інформації.

Поняття інформаційного потоку в обчислювальних системах використовується, в основному при розробці й аналізі алгоритмів обслуговування зовнішніх пристроїв комунікаційними каналами процесора. При цьому використовуються різні моделі обслуговування, однак змістовні аспекти інформації враховуються лише з погляду наявності у потоків даних пріоритетів на обслуговування.

Інформаційні моделі *систем керування і прийняття рішень* одержали поширення відносно недавно. Проблема оцінювання кількості інформації, створюваної автоматичною системою в результаті обробки емпіричних даних, набула особливої актуальності з розвитком автоматизованих систем керування. Існуючі структурні і статистичні оцінки не дозволяють врахувати інтелектуальні витрати на розв'язання задачі керування і значимість одержуваного результату. У зв'язку з цим А. М. Колмогоровим запропоновано вимірювати кількість інформації, що отримується в результаті обробки даних, складністю алгоритму отримання результату. За А. М. Колмогоровим *алгоритмічна інформація* визначається як різниця *алгоритмічних ентропій*

$$A(v:x)=K(x)-K(x/y), \tag{4.32}$$

а алгоритмічна ентропія визначається складністю відтворення об'єкта x:

$$\begin{aligned} K(x) &= \min \cdot l(p_1), \quad G(p_1)=x; \\ K(x/y) &= \min \cdot l(p_2), \quad G(p_2,y)=x, \end{aligned} \tag{4.33}$$

де p_1 – алгоритм одержання результату при відсутності вихідних даних; p_2 – алгоритм одержання результату обробки інформації при вихідних даних y; l – довжина алгоритму; G – деяка обчислювана функція.

Чисельне значення кількості інформації залежить від вибору функції G. Фактично G є сукупністю засобів операційної системи ЕОМ, що виконує алгоритм p. У теоретичних роботах для досягнення єдності одержуваних результатів використовуються кілька мір складності алгоритмів, що виражаються довжиною нормального алгоритму в заданому алфавіті, числом станів машини Тьюрінга й інші. Відомо, що між різними мірами складності існує функціональна залежність.

Модель А. М. Колмогорова зручна при дослідженні одиничних інформаційних процесів обробки даних, але погано узгоджується з моделями масових процесів сприйняття і передавання інформації.