

УДК 655.41

*В. М. Сеньківський, І. В. Піх, Р. О. Коза, Р. Й. Рінецький*

**УДОСКОНАЛЕНА МОДЕЛЬ СТРУКТУРНОЇ ІДЕНТИФІКАЦІЇ СКЛАДНОГО  
КНИЖКОВОГО ВИДАННЯ**

*Запропоновано структурний підхід до створення інформаційної моделі складного книжкового видання. Використано стратифіковане описання вимог, які стосуються правил опрацювання та просторового формування ієрархічних рівнів видання. Розроблено ієрархічну*

модель структури складного видання.

*The structural approach to creation of information model of complex book issuing is offered. The stratified description of the requirements concerning to rules of processing and space creation of hierarchical levels of issuing used. The hierarchical model of structure of complex issuing is developed.*

Розроблення автоматизованої системи проектування книжкових видань [5] – важливої компоненти додрукарської підготовки публікацій – диктує необхідність побудови формалізованої моделі видання як основи інформаційної бази даних системи. З цією метою запропоновано спосіб розв'язання проблеми багаторівневого перетворення тексту й одержання сторінок складних видань [3] у комп'ютерних видавничих системах, що враховує положення, поняття і засоби загальної теорії систем [7 – 10, 12], зокрема одного з її напрямів – теорії ієрархічних багаторівневих систем [8]. При цьому використано стратифікований багаторівневий спосіб описання даних і вимог, які стосуються правил опрацювання та просторового формування ієрархічних рівнів у межах книжкової сторінки.

Основним і визначальним об'єктом перетворення у видавничих системах служить текстова інформація. Поняття «текстова інформація» об'єднує власне текст, формули (математичні, хімічні), таблиці, рисунки, графіки і т.п. Абстрагуючись від змісту тексту, розглядатимемо тільки числові характеристики елементів видання, параметри його формування і стани інформації на різних етапах опрацювання. Сукупність різних відомостей про текстову інформацію назвемо тезаурусом інформаційного забезпечення видавничих систем. У процесі опрацювання текстової інформації тезаурус інформаційного забезпечення змінюється, програмним способом утворюються спеціальні команди для приймання результуючих даних вивідною секцією та наступного одержання друкарських форм.

Застосування багаторівневого аналізу для всієї системи вимагає утворення засобів і методів аналізу на окремих рівнях, кінцевою метою якого є одержання такої сукупності алгоритмів перетворення вхідної текстової інформації, які б забезпечили одержання якісної складальної і, як результат, друкарської форми. Цю глобальну проблему треба розділити на підпроблеми. Спочатку зробимо такий поділ на інформаційному рівні.

Введемо поняття рівня структурної ідентифікації видання, у ролі якого виступатиме логічно і структурно завершена його частина, а саме: символ, слово, рядок, формула, таблиця, ілюстрація, абзац, сторінка, розділ. Очевидно, що між ними існують ієрархічна залежність і взаємозв'язок як за структурою, так і за змістом. Стан сформованості тексту чи готовності інших (нетекстових) елементів на кожному з рівнів у даному випадку суттєвого значення не має. Не вникаючи у внутрішню будову рівнів, розглянемо теоретичні і технологічні вимоги, що пред'являються до локальних перетворень елементів на кожному з них. З усього переліку вимог виділимо ті, які мають відношення до етапу алгоритмізації та програмування задач опрацювання текстової інформації.

Почнемо з найнижчого рівня, яким є символ. Тут опрацювання тексту здійснюється за допомогою різних перетворень, що проводяться над символами. Основними параметрами, які аналізуються при цьому, є код символу, сукупність його ознак, керуючі сигнали або команди, які вказують на характер дій, що стосуються символу. У процесі початкової трансляції символи переводяться в коди внутрішнього алфавіту системи. Додаткові ознаки (буква, голосна, приголосна) в основному використовуються алгоритмами знаходження переносів у слові. До керуючих сигналів (команд), які містять інформацію про майбутній вигляд тексту, належать: «шрифтове виділення», що визначає графічне насичення і нахил рисунка знака; «верхній регістр» – для одержання великих букв, «нижній регістр» – для малих букв; «гарнітура», «кегель» – визначають тип і висоту рисунка знака.

Прикладом системних дій над символами може бути програмна реалізація правил складання, де на основі аналізу кодів символів і їх логічного порівняння приймається рішення про занесення у відповідне місце тексту коду проміжку. Найповніше дії над символами виражені в процесі розділення слів на склади, або пошуку місць переносів у слові. Цей вид опрацювання тексту використовує не тільки код символу, але й різні ознаки, якими він супроводжується. Адже для формалізації правил граматичних переносів в українській мові необхідно відрізнити букви від

«не букв», голосні від приголосних; знати, чи можна відривати букву від решти слова, чи обов'язково залишати її в попередньому рядку; уміти виділяти префікси та інші частини слова.

Наступний рівень – слово. Ця інформаційна структура бере участь у перетвореннях, пов'язаних з формуванням рядків, де вона виступає як складова частина ієрархічно вищого рівня. Слово – основний елемент для реалізації переносів, для чого аналізуються частини слова: буквосполучення, склади, букви. Поділ слів на склади здійснюється на основі граматичних правил орфографії. Додатково використовуються різні методи і методики, які вдосконалюють машинний пошук переносів. До основних відносять метод логічних правил, метод словника, метод поділу слів за звучністю, ймовірнісний метод, матричний метод задання місць переносів [4]. Для одержання потрібної точності переносів слів недостатньо використовувати тільки перераховані методи. Найбільше придатний для формалізованого опису і досить ефективний метод логічних правил доповнюється різними методиками аналізу слова та його частин. Це зазвичай методики виділення слова, що містить тільки букви; визначення довжини слова; пошуку префікса. Їх можна використовувати як сумісно, так і окремо для підготовки, наприклад, комп'ютерних словників.

Опрацювання тексту на рівні рядка вимагає реалізації більш складних вимог, продиктованих відповідною логікою текстових перетворень. Кінцева мета цих перетворень полягає в забезпеченні рівності сумарної ширини деякої множини символів (включно з проміжками між словами і літерами та частиною слова, яка помістилася в рядку, при наявності переносів) величині заданого горизонтального формату [13]. Теоретичні викладки і технологічні обґрунтування щодо правил форматування тексту та обмежень, яких треба при цьому дотримуватись, полягають у наступному [1].

Текстовий рядок послідовно заповнюється словами. Визначається довжина кожного слова, яка додається до поточної довжини рядка. Ця величина після чергового підсумовування порівнюється із заданим горизонтальним форматом:

$$\sum_{i=1}^n \sum_{j=1}^{m_i} a_{ij} + \sum_{i=1}^n (m_i - 1) s_{\min} + (n-1) t_{\min} \leq F, \quad (1)$$

де  $n, m_i$  – відповідно кількість слів і символів у кожному з них;  $a_{ij}$  – ширина  $j$ -го символу в  $i$ -му слові;  $t_{\min}$ ,  $s_{\min}$  – мінімальні значення пробілів між словами та символами;  $F$  – формат рядка.

Доки у формулі (1) зберігається нерівність, накопичення слів продовжується. При досягненні рівності проводиться аналіз останнього слова. Якщо воно не є прийменником, що містить одну букву, або скороченням типу «с.», «м.», «о.» і т. п., то формування рядка завершується. Інакше останнє слово переходить у наступний рядок, а поточний вважається сформованим, якщо його кінець попадає в так звану зону форматування, і при цьому виконується така умова:

$$F - \left( \sum_{i=1}^{N+1} \sum_{j=1}^{m_i} a_{ij} + \sum_{i=1}^n (m_i - 1) s_{\min} + N t_{\min} \right) \leq Z, \quad (2)$$

$$\text{де } N = n - 1, Z = \sum_{i=1}^n (m_i - 1) (s_{\max} - s_{\min}) + N (t_{\max} - t_{\min}).$$

Вираз для характеризує область закінчення рядка або зону форматування. Як бачимо, її величина залежить від кількості слів і символів у рядку та параметрів пробілу між ними. Якщо умова (2) не виконується, тобто рядок не досягає зони форматування, доведення його до потрібної довжини здійснюється за рахунок збільшення пробілів.

Коли після додавання чергового слова довжина рядка перевищує формат, тобто

$$\sum_{i=1}^n \sum_{j=1}^{m_i} a_{ij} + \sum_{i=1}^n (m_i - 1) s_{\min} + \sum_{p=1}^r a_p + (r-1) s_{\min} + n t_{\min} > F, \quad (3)$$

(де  $a_p$  – ширина  $p$ -го символу слова, що не помістилося в рядку), останнє слово ділиться на склади за спеціальним алгоритмом. Після того один або декілька складів опрацьованого слова можуть залишитися в даному рядку, якщо його стан після завершення форматування відповідатиме такій нерівності:

$$\sum_{i=1}^n \sum_{j=1}^{m_i} a_{ij} + \sum_{i=1}^n (m_i - 1) s_{\min} + \sum_{k=1}^s a_k + (s - 1) s_{\min} + n t_{\min} + t_{def} \leq F, \quad (4)$$

де  $a_k$  – ширина  $k$ -го символу частини слова ( $s < r$ );  $t_{def}$  – ширина дефісу.

Якщо перший склад не вміщається в рядку, то останнє слово повністю переноситься в наступний рядок.

Основні вимоги до правил формування рядків викладені у відповідних технологічних інструкціях. Формалізація цих вимог являє собою досить складний процес і реалізується одночасно з проектуванням алгоритмів. Наведені вище формули дають тільки загальне уявлення про можливі напрями застосування аналітичних моделей для формалізованого описання та розв'язання задачі формування текстових рядків.

Наступний рівень видання – формула – за структурою та призначенням близький до попереднього рівня – рядка, однак за суттю тут розв'язується багатофункціональна задача автоматизованого опрацювання та просторового двовимірного компонування математичних формул [11]. При цьому з використанням засобів математичної лінгвістики здійснюються структурування та стратифіковане описання складових частин формул, програмним способом проводиться розрахунок кеглів і присвоєння шрифтів, визначаються межі дробових та індексних виразів. Суттєвим в цій задачі є те, що початково формула за допомогою спеціальної метамови подається у вигляді умовного текстового рядка, який складається зі спеціальних команд і структурно неподільних її частин. Дальший етап – автоматичне розміщення елементів формули на тих місцях, які відповідають їх суті та призначенню. Це один з можливих та ефективних підходів, однак у більшості КВС компонування математичних формул здійснюється в автоматизованому режимі; якість результату в цьому випадку залежить від кваліфікації та рівня підготовки користувача.

Подібним способом, а саме в «ручному режимі» готуються такі специфічні елементи складності видань, як таблиці. Тут ще більше великої ваги набуває суб'єктивний фактор. Процес формування таблиць є кропітким, хоч і не дуже складним. Найчастіше проблеми виникають на етапі комп'ютерного верстання сторінок, що містять таблиці.

На наступному інформаційному рівні, яким є абзац, вимоги стосуються групи рядків. Обмеження накладаються на довжину кінцевого рядка абзацу, величину коридорів та їх просторове розміщення, кількість розміщених підряд переносів. Звертається увага на щільність тексту та її рівномірний розподіл у межах абзацу. Суттєвою і непростюю для реалізації є вимога ліквідації останнього рядка сторінки, який одночасно є першим рядком абзацу. Аналогічною є заборона починати сторінку останнім рядком абзацу. Це так звані «висячі рядки», і завдання недопущення появи їх у виданні пов'язане з втягненням або витягненням рядків, що приводить до зміни щільності тексту.

Від повноти виконання вимог на такому важливому інформаційному рівні, яким є сторінка, фактично залежить результат роботи всієї системи. Оскільки досконалість видання оцінюється якістю верстки, то ступінь виконання вимог завершування рубрик, формул, таблиць, ілюстрацій, виносок, віршів, норм, сигнатур, колонтитулів свідчить про якість алгоритмів опрацювання інформації на цьому рівні. Спрощений варіант основної вимоги вказує на необхідність доведення вертикального розміру сторінки до заданого формату. Математична модель стану сторінки після її формування характеризується співвідношеннями [2]: для мінімальних значень вертикальних проміжків між блоками відповідно для максимальних значень цих проміжків

$$\sum_{r=1}^k h_r + \sum_{i=1}^n \sum_{j=1}^m (t_{ij_{\min}} + \Delta h_{ij_{\min}}) = H; \quad (5)$$

$$\sum_{r=1}^{k+1} h_r + \sum_{i=1}^n \sum_{j=1}^m (t_{ij_{\max}} + \Delta h_{ij_{\max}}) = H, \quad (6)$$

де  $H$  – вертикальний формат сторінки;  $h_r$  – висота  $r$ -го блока;  $k$  – кількість блоків, що помістилися на сторінці;  $t_{ij_{\min}}, t_{ij_{\max}}$  – відповідно мінімальне та максимальне значення проміжку;  $\Delta h_{ij_{\min}}, \Delta h_{ij_{\max}}$  – відповідно приріст мінімального та частка зменшення максимального значень для проміжків даного типу.

Вирази (5) і (6) свідчать про формальне завершення процесу верстання складної книжкової сторінки згідно із заданим алгоритмом та вертикальних числових розмірів блоків і проміжків між ними. Завершення процесу формування сторінки вимагає додаткових дій, що є предметом додаткового дослідження.

Вимоги до формування спускових і кінцевих сторінок, а також розгортки є спільними як до розглядуваного рівня, так і для наступного – розділу. Інші спеціальні вимоги, що стосуються цього рівня, не передбачаються.

Ілюстрації прямого зв'язку із введенням тексту не мають. До ілюстрацій, з точки зору видавничих систем, належать: креслення, рисунки, живопис, фотографії, репродукції з картини [6]. Цей елемент видань готується автономно за допомогою спеціальних пакетів комп'ютерної графіки і розміщується на сторінці художником-дизайнером у процесі верстання тексту, ілюстрацій та інших елементів оформлення. До завдань коригування на цьому рівні відносяться вибір варіантів розміщення ілюстрацій на сторінці, растрування, кольороподіл, суміщення кольорів (трепінг), підбір форматів графічних файлів, коректура зображення. Найбільш популярні графічні пакети для роботи із зображеннями – Adobe Photoshop і CorelDRAW.

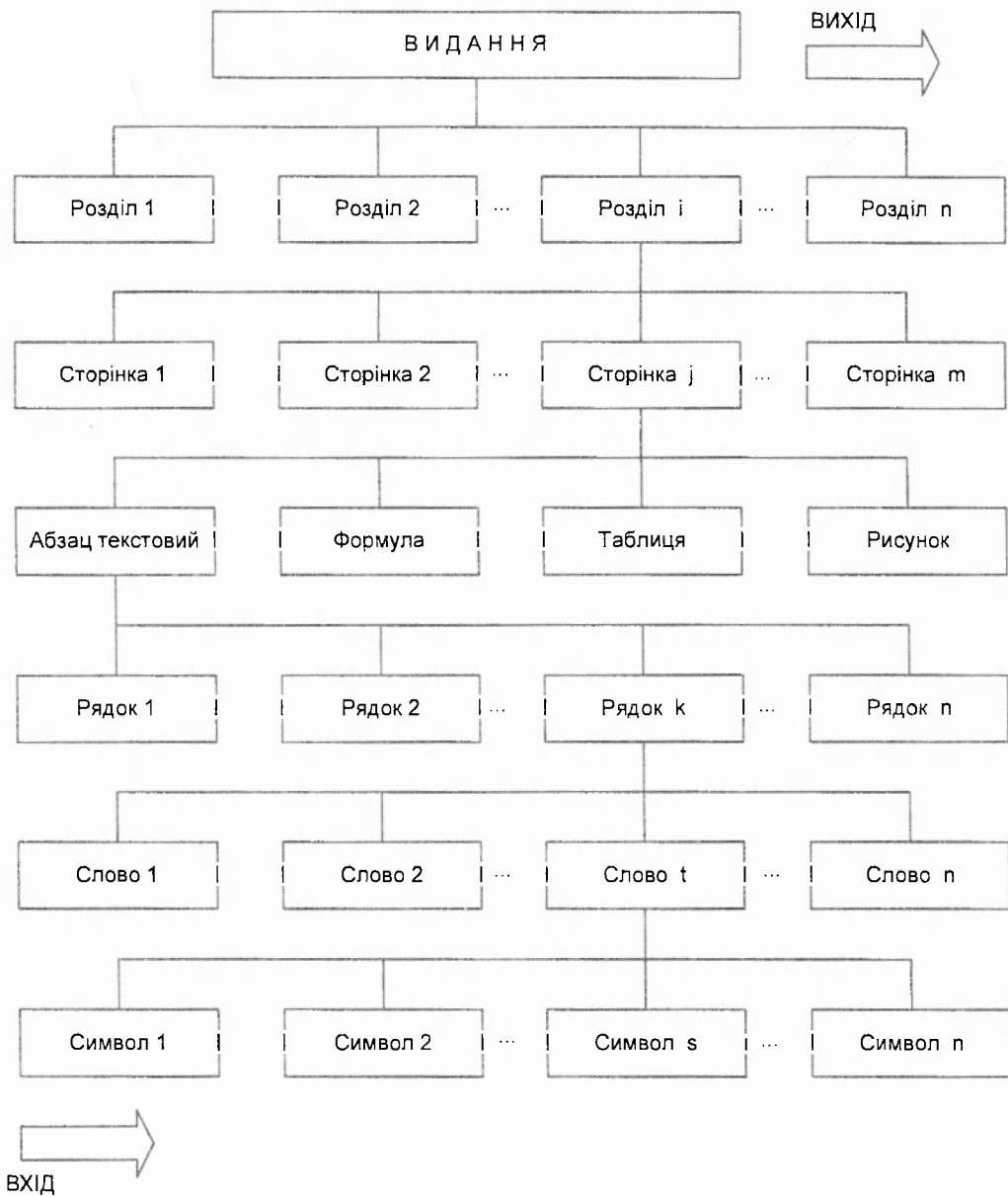
Для рівня видання визначається приведення загального об'єму тексту до заданого; кратність зверстаних сторінок друкарському аркушеві; утворення шаблонів, що відображають структуру книжки; вибір шрифтів і стилів абзаців; задання палітри кольорів.

Результат виконаного аналізу подамо у вигляді узагальненої моделі структури книжкового видання (див. рисунок).

При цьому вважатимемо, що гіпотетичне видання поділене на розділи; крім звичайного тексту (у моделі «абзац текстовий»), воно може містити формули (математичні чи хімічні), таблиці, рисунки (на схемі вони об'єднані загальною назвою «ілюстрації»). Така модель, записана в процесі введення видання в комп'ютер кодами символів, що відображають її смислове навантаження, є по суті основною динамічною частиною загального інформаційного середовища використовуваної комп'ютерної видавничої системи. Вона уможливує опрацювання, зберігання числової моделі видання, передавання (при наявності технічних засобів) за допомогою каналів зв'язку, виведення на електронний носій та остаточне оформлення у вигляді друкованої публікації. Статична частина кожної видавничої системи умовно незмінна й містить широкий набір інформаційних і так званих інструментальних засобів для опрацювання видань довільної складності.

Вхід у змінну частину інформаційного середовища видавничої системи здійснюється на рівні символів множини даних видання, вихід – на рівні цифрової моделі готового видання. Очевидно, що кожний рівень характеризується символічним поданням, тому сформована текстова інформація має символічну структуру, а її стратифікований опис у системі здійснюється засобами того рівня, на якому зафіксовано в даний момент стан видання. Аналогічне твердження стосується інших інформаційних рівнів.

Таким чином, аналізуючи стратифіковане зображення текстової та графічної інформації, абстрагуючись від змісту тексту і керуючись тільки його структурними особливостями й логікою поетапного перетворення ієрархічного інформаційного рівня, приходимо до узагальненої моделі, яка відображає схему побудови видання та взаємозалежність рівнів між собою, уможливує її використання для розроблення автоматизованої системи проектування книжкових видань. Крім того, смисловий аналіз заданих рівнів визначає найважливіші вимоги до подальшого моделювання та алгоритмізації процесу комп'ютерної підготовки видання і створення відповідного програмного забезпечення.



### Інформаційна ієрархічна модель структури видання

Підтверджуються відомі фундаментальні теоретичні засади, а саме: чітко виявляється вертикальна супідрядність елементів; керуючі дії направлені зверху вниз і визначаються рівнями (страгами) з вищим порядковим номером; повнота опису та його деталізація спадають при переміщенні на верхні ешелони; забезпечується повна взаємозалежність і взаємообумовленість дій між рівнями. Крім прямих керуючих дій кожного верхнього рівня на наступний нижній, модель забезпечує зворотний зв'язок, спрямований знизу вгору. Маємо замкнутий інформаційний цикл, характерний для багаторівневих ієрархічних інформаційних структур і систем.

Послідовно описуючи способи перетворення і стани текстової ієрархічної структури на окремих страгах, приходимо до розуміння й обґрунтованого уявлення про загальну структуру системи алгоритмів (програм) комп'ютерної видавничої системи. Має місце взаємно-однозначне відображення інформаційних і програмних рівнів. Ієрархічна багаторівнева структура програмного забезпечення систем комп'ютерного підготування видань слідує з такої ж структури об'єкта перетворення – текстової інформації.

1. Андріїв І.В., Сеньківський В.М. Обґрунтування основних параметрів комп'ютерного формування тексту // Наукові записки. Львів / УАД, 2001. Вип. 3. С. 52–57.
2. Андріїв І.В., Піх І.В., Сеньківський В.М. Оптимізація процесу верстання сторінок книжкових видань з використанням графів // Наукові записки. Львів / УАД, 2003. Вип. 6. С. 79–84.
3. Андріїв І.В., Піх І.В., Сеньківський В.М. Критерії оцінки складності книжкових видань в умовах комп'ютерних технологій // Наукові записки. Львів / УАД, 2002. Вип. 5. С. 71–74.
4. Берлин А.С. Системы программирования набора. М., 1971.
5. Козак Р.О. Узагальнена модель системи проектування книжкових видань // Наукові записки. Львів / УАД, 2004. Вип. 7. С. 63–66.
7. Месарович М., Такахара И. Общая теория систем: математические основы. М., 1978.
6. Комолова Н.В. Компьютерная верстка и дизайн. СПб., 2003.
8. Месарович М., Мако Д., Такахара И. Теория иерархических многоуровневых систем. М., 1973.
9. Молчанов А.А. Моделирование и проектирование сложных систем. К., 1988.
10. Пономаренко О.І., Пономаренко В.О. Системні методи в економіці, менеджменті та бізнесі: Навч. посіб. К., 1995.
11. Ріпецький Р.Й. Технологія автоматизованого компонування текстів математичних формул в процесі випуску складних видань: Автореф. дис... канд. техн. наук / УАД. Львів, 2000.
12. Ситник В.Ф., Писаревська Т.А., Єрьоміна Н.В., Краєва О.С. Основи інформаційних систем: Навч. посіб. / За ред. В.Ф.Ситника. К., 1997.
13. Сеньківський В.М. Методологія проектування систем комп'ютерного підготування видань: Дис. ...док. техн. наук: 05–05–01. Львів, 1996.