

ПАРАМЕТРИ КОНСТРУКЦІЇ КНИЖКИ ТА ІНФОРМАЦІЙНОЇ КОМПОНЕНТИ

Конструкція книжки є однією з ключових компонент книжкового видання (КВ), яка визначає всі основні технічні параметри, споживчі параметри на цілий ряд інших параметрів, що характеризують КВ. В рамках конструкції книжки реалізуються не тільки основна інформаційна компонента, а більшість допоміжних інформаційних компонент. Тому, в першу чергу розглянемо допоміжні інформаційні компоненти і методи визначення їх семантичної узгодженості з базовою інформаційною компонентою, яка вміщає основне інформаційне наповнення КВ. Будемо виходити з того, що базова інформаційна компонента (ВІК) представлена в цифровій формі, що на початковому стані проектування КВ завжди має місце, оскільки рукописи переважно виконуються на електронних засобах набору тексту, або у випадку, коли останні представляються у вигляді рукописного матеріалу, то останній переводиться в цифрову форму. Для того, щоб можна було проводити семантичний аналіз допоміжних інформаційних компонент (DІК), необхідно ввести параметри, якими можна було скористатися при проведенні семантичного аналізу. Одним з таких параметрів є параметр, який характеризує частоту використання слів в деякому тексті [1]. Визначення цього параметру будемо проводити по відношенню до ключових слів, що використовуються в тексті ВІК, до яких не відносяться такі слова, як займенники, сполучники і т.д. Відповідний параметр будемо називати величиною значимості слова для даного ВІК, і будемо його позначати символом $\eta(\sum i)$, де i слово. Розширимо цей підхід на фрази, що використовуються в ВІК. Аналогічно, будемо визначати значимість фраз для ВІК, яку будемо позначити $\eta(X_i)$, де X_i окрема фраза. Реалізації автоматизованої системи визначення $\eta(\sum i)$ для тексту Ψ_i ВІК не представляється проблематичною. Визначення $\eta(X_i)$ є більш складним, оскільки в тексті Ψ_i необхідно розпізнавати окремі фрази, але ця проблема також може бути розв'язана, якщо використати правила граматики відповідної мови, що зв'язані із способами формування фраз [2].

Завдяки процедурам визначення $\eta(\sum i)$ і $\eta(X_i)$ в рамках системи проектування можна сформувати словниковий запас відповідної ВІК та фразеологічний словник ВІК. На основі таких словників можна реалізувати різні методи визначення семантичної узгодженості між окремими інформаційними компонентами. Найпростіший спосіб визначення семантичної узгодженості полягає у наступному. Для цього введемо ряд визначень на основі яких будемо формувати параметри, що характеризують в певній мірі семантику окремих інформаційних компонент.

Визначення 1.1. Семантичною активністю деякого слова $\sum i$ в текстовому відображенні інформаційної компоненти будемо називати відношення параметру $\eta(\sum i)$ до кількості всіх базових слів, що використовуються у даному тексті Ψi .

Формально це визначення запишемо співвідношенням

$$S^A(\sum i, K_i) = N(K_i) / \eta(\sum i), \quad (1)$$

де $N(K_i)$ – кількість базових слів в інформаційній компоненті K_i типу ДІК, $\eta(\sum i)$ – параметр, що визначає значимість слова $\sum i$ в K_i типу ДІК. Очевидно, що $S^A(\sum i, K_i)$ не може використовуватися для аналізу семантичної узгодженості всіх типів ДІК, що використовуються КВ. Цей параметр може використовуватись у випадку аналізу семантичної узгодженості з ВІК таких компонент, як анотації, як рецензії, як вступ. Семантична узгодженість змісту з ВІК розглядається на рівні кожного окремого пункту, який здійснює поділ ВІК, з текстовим фрагментом, який цей пункт визначає. Очевидно, що по відношенню до художньої літератури такий спосіб визначення семантичної узгодженості між ДІК типу змісту і ВІК може виявитися не ефективним. Це може обумовлюватися тим, що в художній літературі є допустимими різні текстові відповідно, семантичні відхилення, що обумовлюються використанням гіпербол, алегорії, синонімів і т.д. В цьому випадку, існує можливість проводити семантичний аналіз, але для його проведення необхідно сформувати словники нестандартних, або нетипових інтерпретацій окремих слів та фраз, що є більш складним, оскільки його формування потребує більш глибокого використання семантики предметної області, що описується у відповідному художньому творі. Тому, в більшості випадків, коли мова йде про оцінку якості КВ, то в першу чергу будемо її проводити по відношенню до книжок, що складають клас підручників, наукових монографій і всіх тих видань використання яких є в певній мірі обов'язковим для того, чи іншого контингенту користувачів.

Виходячи з визначення 1.1, можна ввести наступний критерій семантичної узгодженості двох інформаційних компонентів.

Критерій 1.1 Міра семантичної узгодженості задається діапазоном значень різниці між семантичними розподілами S^R двох компонент узгодженість між якими необхідно визначити.

Формально цей критерій можна описати співвідношенням:

$$K^{su} = |S^R(\Psi i) - S^R(\Psi j)|,$$

де Ψi – текстове відображення порівнюваних інформаційних компонент. Семантичний розподіл інформаційної компоненти Ψi визначається способом що ґрунтується на наступних визначеннях.

Визначення 1.2. Семантичний розподілом $S^R(\Psi i)$ текстового фрагменту Ψi називається функція, що описує зміну величини семантичної активності слів тексту по відношенню до розгортки тексту Ψi .

Визначення 1.3. Розгорткою тексту Ψi є представлення його у вигляді однієї строчки, яка відповідає натуральній по свідомості його зчитування, або його запису які приймаються співпадаючими, позначимо його $Y(\Psi i)$.

Для перевірки семантичної узгодженості виробляються тільки ті текстові компоненти, в яких кількість спільних для двох компонент слів в процентному відношенні відрізняються між собою не більше ніж на задану величину. Таку величину будемо називати мірою подібності двох ІК, і будемо її позначати δ (ІК_i: ІК_j) %. В даному випадку процентне відношення береться для того щоб уникнути різниці в кількості однакових слів, яка обумовлюється різними розмірами ІК_i. Оскільки семантичний розподіл представляє собою функцію, яка формується експериментально і визначена на інтервалі довженною $Y(\Psi_i)$, який залежить від розмірів фрагменту, то для визначення $S^R(\Psi_i)$ необхідно нормалізувати цей параметр таким чином, щоб він став не залежним від інтервалу $Y(\Psi_i)$. Точками координат на інтервалах $Y(\Psi_i)$, чи $Y(\Psi_j)$, будуть окремі слова. В силу умови, що визначається обмеженням δ (ІК_i: ІК_j), на розподілі $Y(\Psi_i)$ визначаємо точки \sum_i , для яких $\sum_{ik} = \sum_{jg}$, де $\sum_{ik} \in \Psi_i$, а $\sum_{jg} \in \Psi_j$. По відношенню до цих точок виділяємо $\Delta_i(\Psi_i)$ на $Y(\Psi_i)$ такі, щоб $dk[Y(\Psi_i)] = di[Y(\Psi_j)]$, де $di[Y(\Psi_i)]$ довжина розгортки фрагменту тексту Ψ_i , при цьому має місце співвідношення $d[Y(\Psi_i)] > d[Y(\Psi_j)]$. Довжина фрагменту тексту $di[Y(\Psi_i)]$ визначається кількістю слів у відповідній розгортці тексту Ψ_i . При визначенні довжини приймаються тільки ті слова, для яких визначається їх семантична активність. У встановленому таким чином інтервалі обчислюється середнє значення семантичної активності, яка приймається як активність відповідного інтервалу. Обчислення реалізується у відповідності із співвідношенням:

$$S^A [Y(\Psi_{ik})] = [\sum_{j=1}^{gi} S^A(\sum_{jj})] / gi, \quad (2)$$

де $S^A(\sum_{jj})$ – семантична активність слова \sum_{jj} , що визначається у відповідності із співвідношенням (1.1.), gi – довжина текстового фрагменту К.

Якщо за межами інтервалу $d[Y(\Psi_{ij})]$ в Ψ_i існує \sum_i , для якого $S^A(\sum_i) = \max$, то на основі точки в $Y(\Psi_i)$, що визначається відповідним словом, знову будеться інтервал $d[Y(\Psi_j)]$ і проводиться обчислення $S^A(\sum_e)$ у відповідності із співвідношенням (1.2). Отримані значення $S^A[Y(\Psi_{ik})]$ використовуються для обчислення величини семантичного розподілу у відповідності з наступним співвідношенням:

$$S^R(\Psi_i) = \sum_{k=1}^m SA[Y(\Psi_k)] / m,$$

де m – кількість фрагментів тексту, що були виділені на $Y(\Psi_i)$. Семантичний розподіл для ІК_j у якого $Y(\Psi_j) < Y(\Psi_i)$, визначається у відповідності із співвідношенням:

$$S^R(\Psi_j) = S^A [Y(\Psi_j)].$$

Якщо в процесі аналізу ІК виявиться, що пари ІК є семантично узгодженими однією з яких в більшості випадків є ВІК, то параметри P^{RZ} , P^{PZ} , P^{DZ} можуть використовуватися для визначення якості інформаційної частини не тільки для ІК типу змісту, а для всіх інших ІК. Завдяки семантичному аналізу стає можливим розв'язувати обернену задачу, яка полягає у тому, що в процесі проектування на етапі оцінки параметрів інформаційної частини

можна реалізувати процеси модифікації системи інформаційних компонент зберігаючи при цьому необхідний рівень семантичної узгодженості. Практично це означає, що у відповідності до вимаганих значень параметрів інформаційної частини може мінятися структура ВІК, розширятися анотація, зменшуватися, або змінюватися рекомендації. Для реалізації безпосередніх алгоритмів, що спричиняють зміну показників якості, необхідно більш глибоко проводити семантичний аналіз, завдяки якому появилась би можливість шляхом розширення ДІК та зміни їх параметрів міняти на потрібні величини характеристик якості ВІК. Слід відмітити, що такі методи семантичного аналізу не можуть сформувати способи зміни інформаційного наповнення базової інформаційної компоненти [3].

Розглянуті вище допоміжні інформаційні компоненти відносяться до типу розширених компонент, оскільки вони складаються більш ніж з одної фрази, або одного речення. Завдяки цьому існує можливість проводити семантичний аналіз ДІК цього типу сумісно з ВІК, завдяки чому появляється можливість змінювати значення параметрів, що характеризують якість інформаційної системи КВ в цілому. Крім розширених ДІК в КВ використовуються ДІК сигнального типу, основною відмінністю яких від ДІК розширеного типу є їх розмір. Якщо розміри ДІК розширеного типу повинні становити величини, що визначаються певною кількістю речень, чи фраз, мінімальна допустима кількість яких встановлюється у вигляді деякого порогового значення, то ДІК сигнального типу складаються з кількості фраз та речень, яка нижча вибраного порогу. Переважно, ДІК сигнального типу складається з одного чи двох речень. Прикладом ІК сигнального типу може служити назва інформаційного наповнення, яка розміщується на титульному листку та обкладинці конструкції книги. Іншим прикладом такого типу ДІК можуть служити дані про автора, дані про видавництво і т.д. Наступною особливістю, якого відрізняється ДІК сигнального типу від ДІК розширеного типу, являється форма, або спосіб її представлення. Під формою, в даному випадку, розуміються розмір та тип шрифту, яким реалізується її текстове відображення. Кількість відмінностей у способах відображення ДІК сигнального типу спосіб реалізації такого відображення може буди тісно зв'язаний з певними, переважно, важливими семантичними значеннями, інформації, що розміщується в ВК.

Конструкція книжки є базовою компонентою КВ, оскільки, завдяки конструкції вдається створити фізичний образ книжки. Слід відмітити, що в традиційних конструкціях, які використовуються в поліграфічній промисловості, використовуються найпростіші рішення по створенню основних вузлів конструкції. До таких основних вузлів можна віднести:

- обкладинка книжки,
- вузол кріплення блоку книжки,
- вузол з'єднання блоку книжки з обкладинкою та інші.

В сучасному видавництві книжок проекти конструкції книжки реалізуються на основі певних критеріїв, кожний з яких має певний

пріоритет. Загально прийнята концепція проектування книжки в якості першочергового критерію пропонує вибирати вартість відповідного продукту. Розвиток електронних технологій, які направлені на надання можливостей, на які орієнтована книжка, дозволяє їм успішно конкурувати з КВ, яке розглядається лише в аспекті джерела інформації, в першу чергу, по відношенню до такого параметра як ціна. Тому існуюча тенденція у наданні пріоритету параметру вартості КВ є помилковим. Для того, щоб можна було конкурувати з відповідними електронними виробами, необхідно в якості параметрів і, відповідно, критеріїв вибору способу проектування КВ, вибирати ті параметри, які є характерними для продукції типу КВ і, на сьогоднішній день, є відсутніми у відповідних електронних приладах, що орієнтована на надання певної інформації. До відповідних факторів можна віднести наступні:

- оригінальність конструкції книжки,
- графічне оформлення конструкції книжки,
- інформаційне, текстове замовлення конструкції.
- використання нетрадиційних компонент в конструкції книжки, що доповнюють інформаційне наповнення відповідної книжки і т.д.

Впровадження перерахованих факторів в конструкцію книжки сприяє формуванню цієї конструкції як виробу який має власну художню цінність. Як правило, предмети, що складають певну художню цінність в різних областях творчості, являються одиничними екземплярами. У випадку книжкової продукції кількість відповідних екземплярів художньої творчості повинна відповідати, як мінімум, кількості екземплярів тиражу відповідної книжки. В цьому випадку, вводиться уявлення про міру оригінальності відповідної конструкції книжки. Така міра оригінальності може мати дві компоненти:

- оригінальність авторська,
- оригінальність тиражування.

Під оригінальністю тиражування розуміється міра відхилень величин значень параметрів, що характеризують параметри всіх компонент конструкції книжки. При цьому існують діапазони відхилень, які є допустимими з точки зору признання відповідного товару, як товару оригінального. Відповідні допустимі відхилення визначаються можливостями технологічних процесів, які необхідно використовувати при виробництві певного типу конструкції книги. Таким чином, міра оригінальності тиражування визначається технологічним обладнанням, яке повинно використовуватися при виготовленні конструкції книжки, що представляє собою мистецький твір. Це означає, що не може будь яке поліграфічне виробництво продукувати конструкції книжки, які представляють собою художній витвір. Подібна ситуація має місце в сучасній поліграфії, але стосується лише міри якості продукованих конструкцій книжок. Можна стверджувати, що міра оригінальності є наступним рівнем уявлень про якість відповідного товару, яким, в даному випадку є конструкція книжки.

Оригінальність авторська розглядається тільки в межах виконання умов забезпечення оригінальності тиражування. Основою відмінності оригінальності авторської від оригінальності тиражування є те, що в першому випадку розглядаються відмінності виключно семантичного характеру. Це є особливо актуальним для компонент конструкції книжки, в яких семантика неоднозначно, або точність її відображення має певну величину похибок. У зв'язку з цим, одним з показників міри авторської оригінальності є степінь однозначності і, відповідно, міра точності опису автором тієї, або іншої семантичної суті у компонентах конструкції книжки, яку використовує автор відповідної компоненти. З цього виходить, що чим більш точно і більш повне та однозначне відтворення деякої семантичної суті, тим вище якість авторської оригінальності відповідної компоненти. У випадку продукування книжок проблема авторської оригінальності, а саме, проблема забезпечення найбільш високої якості авторської оригінальності є більш складною ніж проблема авторської оригінальності у інших видах творчої діяльності, наприклад у таких видах творчої діяльності як творчість художників, творчість музикантів, чи творчість письменників. В останніх випадках у авторів існує можливість використання більш широких діапазонів семантичної інтерпретації предмету або теми по відношенню до якої реалізується творчий продукт. Це прийнято називати свободою творчості. У випадку проектування конструкції книжки, як деякого продукту художньої чи мистецької творчості, автор відповідного продукту є пов'язаний рамками інтерпретації певної предметної області, якій присвячено інформаційне наповнення відповідної книжки. Тому, простір для творчості, є дещо обмеженим, або він не є абсолютно вільним. В цьому випадку процеси реалізації продукту творчості є значно складніші, оскільки вони повинні враховувати рамки інтерпретацій, що приведені в інформаційному наповненні і ці рамки не можуть бути не витриманими, при проектуванні конструкції книжки. При створенні конструкції книжки, творчість можна використовувати лише в межах неоднозначності інтерпретації предметної області, яку описує інформаційна частина, або інформаційне наповнення відповідної книжки. Тим не менше, неоднозначність інтерпретації викладеного матеріалу в інформаційному наповненні не може перевищувати певної міри, оскільки, в протилежному випадку може виявитися допустимим таке розширення та уточнення інтерпретації, яке може привести до суттєвої модифікації початкового варіанту семантичної суті інформаційного наповнення.

Приведений аналіз може бути в певній мірі формалізованим завдяки використанню семантичних параметрів, що допускають їх числову оцінку. Завдяки такому аналізу можна отримати методи числової оцінки таких параметрів як міра оригінальності авторської та міра оригінальності тиражування, які можуть використовуватися для аналізу якості книжкової продукції

Графічне оформлення конструкції книжки полягає у розміщенні на елемент конструкції графічних образів, використання кольорового покриття поверхх книжки та реалізацію різних способів художнього оздоблення

відповідної конструкції.

Графічні образи відрізняються від інших, приведених вище графічних компонент тим, що вони в найбільшій мірі пов'язані семантично з інформаційним наповненням книжки. Це означає, що всі умови еквівалентності семантики інформаційного наповнення та семантики образу мусять виконуватися. До таких умов семантичної еквівалентності відносяться наступні:

- відсутність семантичної суперечності,
- відсутність семантичних конфліктів,
- збереження семантики в границях предметної області інтерпретації,

що визначені в інформаційному наповненні книжки.

Виявлення відсутності семантичної суперечності між інформаційним наповненням (IN) книжки та графічним образом потребує використання досить складних процедур, які в першу чергу пов'язані з формуванням текстового опису відповідного графічного образу. Ця задача може пов'язуватися наступними способами.

Перший спосіб ґрунтується на тому, що графічні образи передбачається створювати на стані проектування конструкції книжки. В цьому випадку, існує можливість однією з умов створення графічного образу тим, або іншим автором, повинно бути формування текстового опису відповідного образу у відповідності з яким, з певною мірою адекватності міг би формуватися відповідний графічний образ.

Другий спосіб ґрунтується на розпізнаванні готового графічного образу і на основі даних про розпізнавання формування текстового опису цього образу. Цей спосіб представляється більш доцільним в силу наступних причин та факторів:

- опис в текстовій формі графічного образу, який необхідно нарисувати, може виявитися не повністю, або не адекватно описувати образ,
- нарисований образ, навіть, якщо він створювався на основі його попереднього опису може представляти собою більш широкую інтерпретацію інформаційного змісту який представлено автором образу,
- приведені вище фактори обумовлюють необхідність реалізувати додаткову процедуру аналізу на семантичну узгодженість образу і тексту, що його описує.

Другий спосіб передбачає використання автоматизованих способів розпізнавання, під яким розуміється формування текстового опису відповідного образу, що суттєво легше реалізувати по відношенню до розпізнавання, яке полягає у ідентифікації образу певному еталону [4].

В рамках цього підходу існує можливість формувати опис образу, який розпізнається і з довільною мірою адекватності формувати його опис. Така міра адекватності визначається на основі аналізу величини семантичної однозначності, яка відповідає текстовим фрагментам, що відносяться до сюжету графічного образу. Крім того, отримані текстові описи можна аналізувати на предмет наявності семантичних аномалій.

Конструкція книги завжди наповнюється текстовими фрагментами, які безпосередньо зв'язані з ІN. Мінімальні текстові наповнення, які є обов'язковими, представляють собою назву ІN, а ідентифікаційні дані автора ІN, а також інші дані, що інформують про видавництво, яке відповідну книжку випродукувало, рік видавництва, офіційні дані видавництва, що свідчать про його уповноваженість до видавничої діяльності.

Якщо розглядати KB, як продукт художньої творчості, то інформаційне наповнення конструкції книжки повинно бути більш повним. В цьому випадку, елементом творчості є не тільки зміст тих фрагментів тексту, що розміщуються в рамках конструкції книжки, а і форма представлення відповідного тексту. Під формою представлення розуміється кольор шрифтів, графіка шрифтів їх місце розміщення і т.д.

1. *Zipf G.K.* Human beharar and the principle of least effort. Adolison – Wesley, Reading MA, (USA) 1949.
2. *Афанасьев А.Н.* Формальные языки и грамматики, Ульяновск: УлГТУ, 1997.-82с.
3. *Шевчук А.В.* Використання теорії формальних граматик для розв'язування задач побудови графічних засобів захисту поліграфічної продукції // Вісник житомирського державного технол. Університету ЖТГІ, №3,2002.
4. *Будько М.М.* Біометрична інформація та ідентифікація особи./ Київ. Укр ІНТЕІ.200.

Поступила 2.02.2010р.

УДК 004.942

Р.А. Бунь^{1,2}, У.І. Матвійчук¹

ГЕОІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПРОСТОРОВОГО АНАЛІЗУ ЕМІСІЙ ТА ПОГЛИНАНЬ ПАРНИКОВИХ ГАЗІВ У ЛІСОВОМУ ГОСПОДАРСТВІ ЗАКАРПАТСЬКОЇ ОБЛАСТІ

Summary. An approach for formation of mathematical models for spatial analysis of emission and sink of greenhouse gases in forestry sector of Zakarpattia region has been presented. The created geoinformation technology gives a possibility of georeferenced database formation, and spatial cadastre calculation for emissions and sinks.

Keywords: mathematical modeling, geoinformation technology, greenhouse gas inventory, forestry sector, Zakarpattia region.

Анотація. Представлено підхід до формування математичних моделей для просторового аналізу процесів емісії та поглинання парникових газів у лісовому

¹ Національний університет "Львівська політехніка"

² Академія інформатики та управління м.Бельско-Бяла, Польща