

### 4.3 Семантичні мережі і гіпертекстові системи

Семантичні технології забезпечують існування певного рівня абстракції над існуючими інформаційними ресурсами. Цей рівень дозволяє здійснювати зв'язок даних, змісту і процесів між різними видами ізольованих структур даних та інформації. Основою семантики є онтології, що призначені для формалізації предметних областей. Аналіз онтологій є складною задачею, оскільки передбачає обробку великих обсягів даних та логічних тверджень, які містяться у онтологіях.

Останнім часом дослідження семантичних інформаційних моделей здійснюється переважно стосовно гіпертекстових структур і гіпермедіа, які є основою Web-систем.

#### 4.3.1 Поняття про семантичні моделі

Семантичні складові статей гіпертекстових систем формують так звані семантичні мережі. Семантичні технології подають знання за допомогою онтологій і забезпечують зв'язки, правила, логіку та умови, що зазначені в онтології. Тому онтології можна вважати основою семантичної мережі.

Онтологія – формальна специфікація інформаційної моделі предметної області. Онтологія складається з класів сутностей предметної області, властивостей цих класів, зв'язків між цими класами і тверджень, побудованих із цих класів, їх властивостей і зв'язків між ними.

Існують різні класифікації онтологій:

1. семантична:

- за рівнем виразності;
- за ступенем формальності;
- за рівнем детальності подання;

2. прагматична:

- за ступенем залежності від конкретної задачі чи предметної області;
- за мовою подання онтологічних знань;
- за предметною областю;
- за метою створення.

Наведемо коротку характеристику кожної класифікації.

#### *Семантична класифікація*

1. За рівнем виразності.

*Глибинні онтології.* Глибинні онтології сильно аксіоматизовані. Мета аксіоматизації – уникнення термінологічної і концептуальної неоднозначностей. Кожна глибинна онтологія може мати поверхневу версію. Багато онтологій предметних областей є глибинними, оскільки повинні підтримувати процеси побудови складних висновків.

*Поверхневі онтології.* Це прості структури примітивів або композиції термінів з відповідними означеннями. Вони слабо аксіоматизовані, оскільки значення терміна, використовуване всередині співтовариства, зазвичай, більш-менш відомо усім його членам.

2. За ступенем формальності.

Ця класифікація схожа з класифікацією за рівнем виразності мови опису онтології, проте не еквівалентна їй.

*Неформальні.* Це онтології, які описуються в документі будь-якою природною мовою. Незважаючи на відсутність формальних правил задання, такі онтології також можуть бути суттєво наповненими, несуперечливими і точними.

*Частково формалізовані.* Такі онтології, хоча і формалізовані, але дуже слабо структуровані.

*Сильно формалізовані.* Онтології для вирішення інженерних завдань. Здають формальну семантику термів (таких, як кількість і одиниця виміру) в дозволених мовою точних і несуперечливих виразах.

3. За рівнем детальності подання.

*Низько деталізовані.* Онтологія може бути побудована на основі термінів і декількох типів зв'язків.

*Високо деталізовані.* Онтологія може містити набагато більше деталей, охоплюючи правила, за якими терміни можуть бути пов'язані між собою.

#### **Прагматична класифікація**

1. За ступенем залежності від конкретної задачі чи предметної області.

*Верхнього рівня.* Такі онтології описують найбільш загальні поняття (простір, час, матерія, об'єкт, подія, дія і т. д.), які не залежать від конкретної проблеми чи області.

*Орієнтовані на предметну область.* У ряді дисциплін наявні стандартні онтології, які можуть використовуватися експертами з предметних областей для спільного використання та анотування інформації у своїй галузі. Наприклад, у галузі медицини створені такі великі стандартні структуровані словники, як SNOMED і семантична мережа Системи уніфікованої медичної мови (Unified Medical Language System). Також з'являються великі спільноцільові онтології, наприклад, онтологія UNSPSC, що надає термінологію товарів і послуг.

*Орієнтовані на завдання.* Це онтології, що використовуються конкретною програмою і містять терміни, які застосовуються при розробці програмного забезпечення, що виконує конкретне завдання. Вони відображають специфіку додатка, але можуть також містити деякі загальні терміни (наприклад, в графічному редакторі будуть і специфічні терміни – палітра, тип заливки, накладання шарів тощо, і загальні – зберегти і завантажити файл).

*Прикладні онтології* описують концепти, які залежать як від онтології завдань, так і від онтології домена. Прикладом може слугувати онтологія для автомобілів, будівельних матеріалів, обчислювальної техніки [114].

### 2. За мовою подання онтологічних знань.

RDF – мова, розроблена в рамках проекту Semantic Web. Основне призначення мови – опис метаданих документів, що розміщуються в мережі Internet. RDF використовує базову модель подання даних “об’єкт – атрибут – значення”, здатну зіграти роль універсальної мови опису семантики ресурсів і зв’язків між ними.

DAML OIL – семантична мова розмічування веб-ресурсів, яка розширює стандарти RDF і RDF Schema за рахунок повніших примітивів моделювання.

OWL (Web Ontology Language) – мова подання онтологій наступного, після DAML OIL покоління. Має більший набір можливостей, ніж XML, RDF, RDF Schema і DAML OIL.

KIF (Knowledge Interchange Format, або формат обміну знаннями) – спеціальна мова, призначена для обміну знаннями між різними комп’ютерними системами. Розроблялась для опису загального формату подання знань, незалежного від конкретних систем.

CycL (мова опису онтології Cyc) – гібридна мова, в якій об’єднані властивості фреймів і логіки предикатів.

OCML (Operational Conceptual Modeling Language) мова підтримує побудову декількох типів конструкцій подання знань. Вона дозволяє задавати специфікацію функцій, зв’язків, класів, екземплярів і правил. Вона також містить механізми для опису онтологій і методів вирішення завдань – основні технології, розроблені у сфері подання знань.

Ontolingua – надає розподілене середовище для спільного перегляду, створення, редагування, зміни і використання онтологій.

### 3. За предметною областю.

Онтологія відбиває такі загальні знання про предметну область, як ієрархія класів понять і семантичні відношення на цих класах. Для кожної предметної області онтології створюються експертами з урахуванням ієрархії понять і традицій області.

### 4. За метою створення.

*Онтології додатка.* Використовуються під час виконання конкретного додатка, що використовується в роботі блока побудови міркувань. Онтології додатка можуть також описувати конкретні світи (семантичні описи, бази знань, метадані, семантичні метадані чи просто екземпляри).

*Посилальні онтології.* Використовуються під час розробки додатків, для взаємного розуміння і тлумачення між агентами, що належать до різних співтовариств, для встановлення консенсусу між співтовариствами, які потребують введення нового терміна, чи просто для пояснення значення терміна новому учасникові співтовариства.

### 4.3.2 Моделі гіпертекстових систем

Основою Інтернет–ресурсів є гіпертекст (ГТ) – одна з фундаментальних моделей подання знань у текстовій формі. Якщо одновимірний текст – однонаправлений рядок символів, то багатовимірний текст (ГТ) є нелінійним, містить точки розгалуження, що дозволяють здійснювати перегляд в декількох напрямках. Відповідно, гіпертекстова система – система, що складається з специфікації інтерфейсу взаємодії з ГТ, визначає принципи та примітиви ГТ (формалізація точок розгалуження та зв'язків між ними), а також принципи зберігання, доставки та знаходження об'єктів ГТ.

Особливості ГТ:

- текст розділений на фрагменти, що є його семантичними одиницями (sets, сети), між якими встановлюються іменовані зв'язки;
- ГТ можна переглядати, рухаючись різними траєкторіями, утвореними зв'язаними сетами;
- переходи вибираються користувачем.

Отже, гіпертекст – форма організації семантичної інформації, що передбачає її фрагментацію та зв'язування.

У гіпертекстовому документі може бути подано декілька рівнів деталізації матеріалу. Такі документи моделюються деревами чи мережами. У графовій моделі ГТ вершини відповідають фрагментам тексту, а ребра – можливим переходам між ними. Кожен шлях на графі показує окрему послідовність перегляду матеріалу.

Згідно з протоколом передачі гіпертексту Hypertext Transport Protocol (HTTP) текст, записаний мовою розмітки гіпертексту Hypertext Markup Language (HTML), є мінімальною одиницею даних, призначеною для міжмашинного обміну. Файл, що містить текст з HTML–розміткою, є гіпертекстовим документом і називається HTML– або WEB–сторінкою. HTML–сторінка містить опис структури документа, до складу якого у вигляді уніфікованого покажчика ресурсу (Uniform Resource Locator – URL) можуть входити посилання на інші гіпертекстові фрагменти. Взаємопов'язана сукупність HTML–сторінок, розташованих на одному WEB–сервері, утворює WEB–сайт.

Гіпертекстовий документ, поданий у HTML, може містити не тільки текстову, але й графічну інформацію, інтерактивні елементи, програмний код мовами програмування (JavaScript, VBScript).

Основні моделі гіпертексту були розроблені і запропоновані у 1990–х роках для забезпечення однозначного та систематизованого опису структури систем гіпермедіа. Залежно від призначення, всі моделі можна розділити на дві основні групи:

- гіпербазові моделі (hyperbase models);
- шарові моделі гіпермедіа (layered hypermedia models).

Залежно від способу подання розглядають такі типи моделей ГТ:

- формалізована модель;
- умовно–типова модель.

Основна задача *гіпербазових моделей* полягає у моделюванні зв'язування даних, а шарових – у моделюванні внутрішньої архітектури систем гіпермедіа.

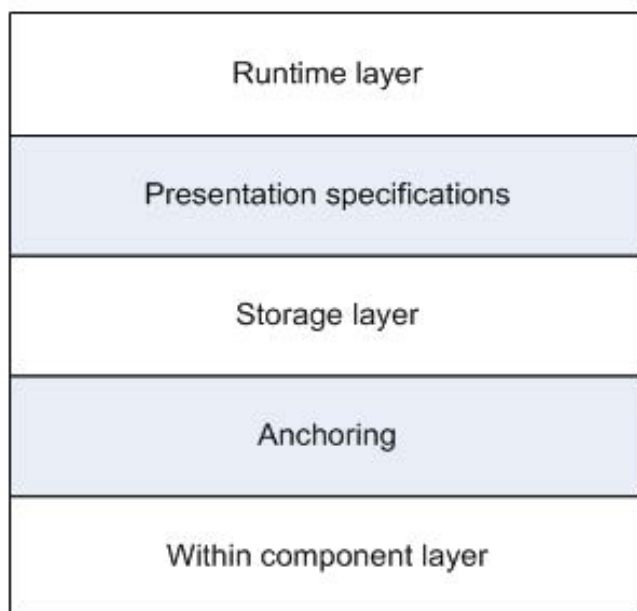
В основі гіпербазових моделей лежать результати досліджень концептуальних моделей даних для опису та оперування інформацією гіпертекстових мереж. Їх основна концепція полягає у поданні інформаційних фрагментів у вигляді вузлів і зв'язків для формалізації відношень між ними.

Шарові *моделі гіпермедіа* являють собою еталонні моделі для розробки систем управління гіпертекстом. Шари в такій моделі відповідають різним вимірам моделювання, а саме: моделі структури і моделі подання даних. Модель Dexter – одна з найвідоміших в даному класі. Ця модель пропонує еталонну архітектуру шляхом запровадження загального словника для порівняння різних гіпертекстових систем та моделей. Як показано на рис. 4.3, в основі моделі є три шари:

– шар зберігання даних (*storage layer*) – основний шар в моделі Dexter. Подає структуру вузол – зв'язок гіпертекстової мережі як “базу даних” компонентів, пов'язаних між собою посиланнями;

– шар часу виконання (*runtime layer*) – належить до динамічних компонентів, а саме: містить засоби перегляду, доступу та модифікації структури мережі;

– внутрішньокomпонентний шар (*within component layer*) – належить до внутрішньої структури та вмісту кожного компонента.



зображення 4.3 – Шарова архітектура за моделлю Dexter

Інтерфейс між шаром часу виконання та шаром зберігання даних називається специфікацією подання (*presentation specification*). Інтерфейс між шаром зберігання даних та внутрішньокomпонентним шаром називається специфікацією прив'язки (*anchoring specification*), пов'язаний з ідентифікацією початкових і кінцевих розташувань посилань у змістовній частині вузлів. Інші моделі, розроблені в цьому напрямку, здебільшого розширюють модель Dexter.

У *форматованій моделі* ГТ інформаційно-довідкова стаття (ІДС) відповідає інформаційному об'єкту, зміст якого характеризується смисловою єдністю та логічною цілісністю. Вона може містити інформацію, подану в різних формах: текст, таблиці, фрагменти програмного коду (макриси, скрипти), інтегровані інтерактивні об'єкти, посилання на подібні об'єкти, що підключаються до ІДС при її завантаженні.

На рис. 4.4 наведено загальний випадок, де як ІДС використовуються і тестові фрагменти і мультимедійні об'єкти.



Рисунок 4.4 – Графова модель гіпермедіа

Елементи ІДС (слова, комірки таблиць, піктограми, зображення, кнопки та ін.) можуть бути гіперпосиланнями.

Для ідентифікації ресурсів у мережі Інтернет використовується URL, що задає адреси відповідних ресурсів. Гіперпосилання, що вказують на фрагменти поточного ресурсу, є локальними, ті, що вказують на зовнішні джерела, мають назву глобальних.

### 4.3.3 Критерії оцінювання гіпертекстових структур

Розглянемо критерії оцінювання гіпертекстових структур на прикладі системи, граф якої зображено на рис. 4.5.

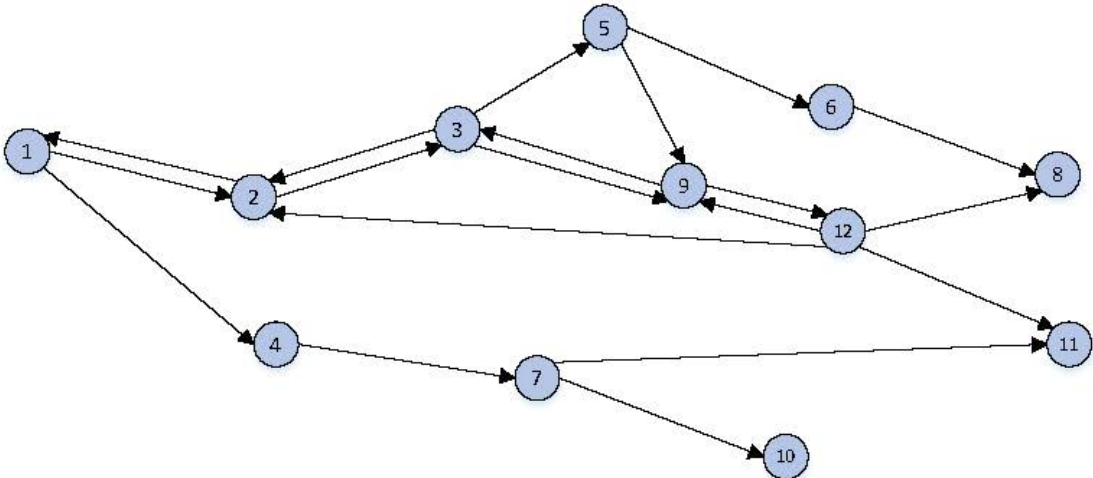


Рисунок 4.5 – Приклад орієнтованого графу

**Індекс інформаційної компактності** відображає ступінь перетину гіпертекстової структури зв'язками.

Високий рівень компактності характеризує такі гіпертекстові структури, в яких на будь-який з інформаційних блоків можна з легкістю потрапити з будь-

якого іншого блока (що, зазвичай, забезпечується численними перехресними зв'язками). Слід зазначити, що надмірно висока компактність може призвести до повної дезорієнтації користувача гіпертекстової системи. В свою чергу, низька інформаційна компактність сприяє випаданню з поля зору багатьох фрагментів гіпертексту чи призводить до втрати досяжності окремих фрагментів.

Індекс інформаційної компактності, визначається так:

$$Cp = \frac{CD_{\max} - CD}{CD_{\max} - CD_{\min}} \quad (4.17)$$

де  $CD_{\max}$  – максимально можливе число кроків, які необхідно зробити за посиланнями, що зв'язують усі вузли гіпертексту;

$CD_{\min}$  – мінімальне можливе число кроків, які зв'язують усі вузли гіпертексту (в тому випадку, коли всі вузли гіпертексту зв'язані зі всіма);

$CD$  – показник шляхів в графі.

Значення індексу інформаційної компактності варіюється в межах  $[0; 1]$  і допускає порівняння систем гіпертекстових документів між собою.

**Частка відсутніх шляхів**  $Q_m$  у графі. Максимальна кількість відсутніх шляхів дорівнює  $n^2 - n$ , мінімальна – 0. Абсолютне значення критерію є незручним. Більш інформативним критерієм є нормована частка відсутніх шляхів

$$K_m = \frac{Q_m}{n^2 - n} \quad (4.18)$$

Значення частки відсутніх шляхів варіюється в межах  $[0; 1]$  і допускає порівняння систем гіпертекстових документів між собою. Для розрахунку коефіцієнта відсутніх шляхів у графі необхідним є попереднє визначення матриці відстаней.

**Цикломатичне число** характеризує відмінність структури графу від деревоподібної і визначається за формулою

$$Cn = m(G) - n(G) + p, \quad (4.19)$$

де  $m(G)$  – число ребер,  $n(G)$  – число вершин,  $p$  – число зв'язних компонент графу.

Цикломатичне число показує найменшу кількість ребер графу, які необхідно видалити, щоб він перетворився у дерево. Для сильнозв'язного графу  $p=1$ . Кістяковим деревом зв'язного графу  $G$  є будь-який його підграф, що містить усі вершини графу  $G$  і є деревом. Якщо  $G$  – зв'язний граф, що містить  $n(G)$  вершин і  $m(G)$  ребер, то кістякове дерево графу  $G$  (якщо воно існує) повинно містити  $n(G) - p$  ребер.

Таким чином, будь-яке кістякове дерево графу  $G$  є результатом видален-



ня з графу  $m(G) - (n(G) - p) = m(G) - n(G) + p$  ребер.

Визначення цикломатичного числа дозволяє знайти число незалежних маршрутів в гіпертекстовій структурі.

**Індекс стратифікації** введений для характеристики лінійності гіпертексту. Фактично індекс стратифікації дозволяє оцінити рівень зв'язності елементів, що стоять на різних рівнях ієрархії.

Обчислення відносного індексу стратифікації базується на операціях з матрицею відстаней орграфу, для побудови якої введемо такі позначення:

- нехай  $G(V, E)$  – орграф;
- $d(u, v)$  – відстань між вершинами  $u$  та  $v$  в орграфі  $G$ ;
- $a_i$  – сума відстаней від  $u_i$  до всіх  $v$ . Таким чином,  $a_i$  – сума значень  $i$ -го рядка матриці відстаней  $D$ ;
- $b_j$  – сума відстаней від  $v_j$  до всіх  $u$  в  $D$ . Таким чином,  $b_j$  – сума кінцевих значень  $j$ -го стовпця матриці відстаней  $D$ ;
- загальна відстань  $(\sum_i \sum_j d_{ij}, d_{ij} \neq \infty)$  в орієнтованому графі  $G$  – сума кінцевих відстаней  $d(v_i, v_j)$ . Тобто,  $(\sum_i \sum_j d_{ij}, d_{ij} \neq \infty)$  – сума всіх кінцевих елементів матриці відстаней.

Для оцінки гіпертексту використовують абсолютну стратифікацію

$$AP = \sum_i |s_i|.$$