

Дядичев В.В., Сафронов К.М.

МЕТОДИ ВІЗУАЛІЗАЦІЇ ІНФОРМАЦІЙНИХ СТРУКТУР

Візуалізація інформації включає в себе візуальне уявлення абстрактних інформаційних просторів і структур для полегшення їх швидкого засвоєння і розуміння. Інформаційні структури включають ієрархії, мережі, багатомірні таблиці, а також колекції текстових документів. Методи, що базуються на візуалізації інформаційних структур знаходять все більше застосування в різних областях науки. Рис. 5. Табл. 1. Джер. 5

Ключові слова: візуалізація, інформація, інформаційна структура, граф, мережа, програмне забезпечення, графічне представлення

Актуальність теми. Візуалізацію інформаційних структур можна охарактеризувати наступним чином: «Візуальне представлення абстрактних просторів інформації та інформаційних структур, що сприяє їх швидкому засвоєнню і розумінню». Це включає в себе процес трансформації абстрактних даних у відповідну візуальну форму. Візуалізація інформації надає безліч ефективних методів для візуалізації інформаційного наповнення, а також для відображення інформаційних структур. При наявності відповідних засобів візуалізації інформаційних структур, можна отримати загальне уявлення дуже великого інформаційного простору, та при необхідності здобути детальну інформацію за допомогою механізмів масштабування та фільтрації.

Люди володіють чудовою здатністю швидкого сприйняття, розпізнавання і пригадування зображень. Зорова система сприйняття людини може швидко і автоматично виявляти закономірності та зміни в розмірі, кольорі, формі, русі, або текстурі зображення. Інтерактивні системи візуалізації інформаційних структур можуть використовувати цю попередню обробку, що дуже важливо для конденсації великих обсягів інформації для її керованих відображень. Візуалізація інформаційних структур прагне представити інформацію візуально, по суті, щоб розвантажити зорову систему сприйняття людини для когнітивної роботи.

Постановка проблеми. Візуалізація інформаційних структур в певній мірі залежить від виду інформації. За аналогією Шнейдермана [1], інформацію можна розділити на такі типи:

- Лінійні: таблиці, програмний код, алфавітні переліки, повідомлення, що впорядковані в хронологічному порядку.
- Ієрархії: деревовидні структури.
- Мережі: загальні графові структури, такі як графи посилання гіпермедіа вузлів, семантичні мережі, веб, і т.д.
- Багатовимірні: атрибути метаданих, такі як тип, розмір, автор, дата зміни, і т.д. Елементи з N атрибутами стають точками в N-мірному просторі.
- Векторні простори: вектор генерується зі змісту елемента, що використовується для представлення його характеристик. Наприклад у згрупованих текстових документах, вектори відображають частоту (ключових) слів, що характеризують кожний документ.
- Просторові: за своєю суттю це 2D або 3D дані, такі як плани поверхів, карти, CAD моделі, комп'ютерна томографія і т.д.

Ієрархічні структури стають все більш поширеними в теперішньому інформаційно-багатому суспільстві як засіб для організації великих обсягів інформації. Прикладами інформаційних ієрархій можуть бути ієрархічні організації працівників у компанії, структура файлів і каталогів, що зберігаються на жорсткому диску комп'ютера. Складність ієрархії зростає, і навіть жорсткий диск звичайного домашнього комп'ютера може містити багато мільйонів файлів. Таким чином, надання візуального уявлення такого, що легко підлягає дослідженню є необхідним для того, щоб зробити великі ієрархічні структури доступними і зрозумілими.

Деревовидна структура має підпорядкованість зверху вниз. Внутрішні вузли мають нащадків, що розташовані нижче по ієрархії, кінцеві вузли є остаточними вузлами, які не мають власних нащадків. У дереві існує єдиний шлях з кореня дерева до кожного кінцевого вузла. Наприклад, в ієрархії файлової системи, внутрішніми вузлами є каталоги і підкаталоги, а кінцевими вузлами є файли. Кожен файл може бути досягнутий від кореня файлової системи, відкривши повний шлях до нього через директорії і піддиректорії. У системі управління документами, внутрішніми вузлами можна було б назвати колекції і під-колекції, а кінцевими вузлами можуть бути документи. У схемі прийняття рішень у якості внутрішніх вузлів виступають вибори, та кінцевими вузлами - остаточні рішення.

Традиційні методи візуалізації, такі як схеми, або діаграми дерев, ідеально підходять для візуалізації малих, компактних інформаційних ієрархій, які можна зробити на одному аркуші паперу або представити на екрані комп'ютера (без прокрутки). Однак, якщо ієрархія є занадто великою, візуалізація не спроможна бути простою і зрозумілою [2].

Багато загальних графів (мереж) можуть бути перетвореними в охоплюючи дерева зі зворотними посиланнями, що дозволить застосувати ієрархічні системи візуалізації до загальних графів та мереж шляхом вилучення і візуалізації сполучного дерева, і подальшого накладення будь-яких зворотних посилань.

Мережі є загальними графічними структурами, що включають набори вузлів, з'єднаних ребрами. У той час як дерево має підпорядкованість зверху вниз і не допускає циклів, граф має довільний зв'язок між вузлами.

Приклади мереж включають в себе веб-сторінки і зв'язки між ними, карту Київського метро зі станціями, що з'єднані за допомогою ліній, або мережі семантичних понять і відносин між ними.

Файли та документи часто мають пов'язані з ними атрибути метаданих, такі як автор, назва, дата зміни, розмір, тип і розширення. Ці атрибути формують N-мірний простір, в рамках якого знаходиться кожен документ. Цей N-мірний простір може бути спроектований з різні 2D або 3D представлення для перегляду і спостереження. Таким чином, замість відображення простого лінійного переліку документів, скажімо, у відповідь на пошуковий запит, атрибути метаданих можуть бути використані для візуалізації наборів документів в багатовимірному просторі.

Розроблено багато різних методів візуалізації інформаційних структур. Ці методи можуть бути класифіковані наступним чином [3]:

Явні та неявні. Явні методи (класичні дерева) малюють по одній лінії для кожної відносини між підпорядкованими елементами. Неявні методи застосовують технологію заповнення простору. Вони відображають відносини, що засновані на спеціальних накресленнях графічних об'єктів, які являють собою ієрархію вузлів [4].

Горизонтальні та радіальні. Горизонтальні компоновки використовують паралельні горизонтальні лінії для організації вузлів. Кожен рядок представляє собою один рівень ієрархії. Всі вузли на тому ж рівні ієрархії, розташовані на тому ж рядку. Радіальні компоновки будуються на основі концентричних кіл. Кожне коло являє собою один рівень ієрархії, тому вузли одного рівня ієрархії розташовуються вздовж одного кола. Кореневий вузол ієрархії знаходиться в центрі концентричних кіл.

Двовимірні та тривимірні. 2D-компоновки відображають ієрархії в двовимірному просторі та будуються на основі 2D елементів. На відміну, 3D-представлення використовують для розташування ієрархії тривимірний простір.

Програмне забезпечення, що служить створенню візуального представлення абстрактних даних, навіть якщо вони сильно відрізняються між собою, слідує процесу генерації, що представлений на рис. 1. Візьмемо вихідні дані в якості відправної точки. Ці дані не обов'язково мають конкретні зв'язки з фізичним простором. Наприклад, вони можуть стосуватися імен людей, цін на споживчі товари, результатів голосування, і так далі. Ці дані рідко доступні в форматі, який підходить для обробки за допомогою автоматичних інструментів і, зокрема, програмного забезпечення візуалізації. Таким чином, похідні дані повинні бути відповідним чином оброблені, перш ніж бути представленими графічно.

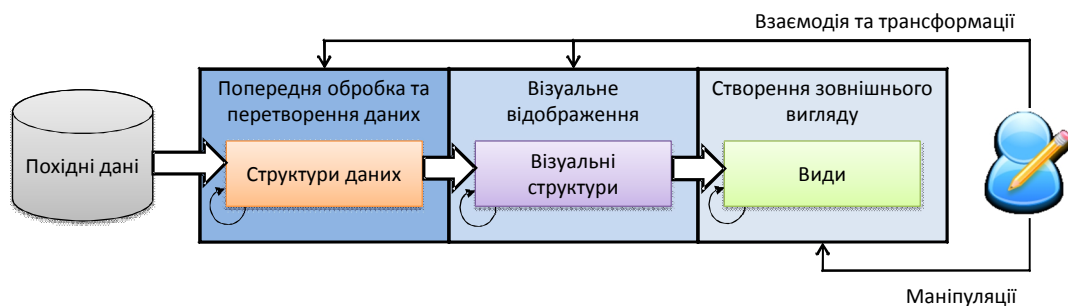


Рис. 1. Процес створення графічного представлення

Створення візуальних об'єктів це процес, який можна змоделювати через послідовність етапів.

1. Попередня обробка та перетворення даних.
2. Візуальне відображення.
3. Створення зовнішнього вигляду.

Опишемо кожен з цих етапів перетворення даних з вихідного формату до створення візуального уявлення.

Попередня обробка та перетворення даних. Термін «первинний» використовується для опису даних, що були отримані з навколишнього світу. Це можуть бути дані, отримані вимірювальними приладами, такі як значення деяких забруднюючих речовин, взятих зі станції моніторингу забруднення під час тестування. Вони також можуть бути створені у результаті розрахунку відповідним програмним забезпеченням, наприклад, дані прогнозу погоди. У кожному разі, ці набори даних дуже рідко мають точну логічну структуру. Для того, щоб обробити ці дані за допомогою програмного забезпечення, необхідно привести їх до організованої логічної структури. Зазвичай для цього типу даних використовуються табличні інформаційні структури - організація даних у таблицю у форматі, що є відповідним для програмного забезпечення, яке має їх обробляти. Іноді вхідні дані, що містяться в одній або декількох базах даних і, отже, вже доступні в електронному форматі з чіткою структурою. У цьому випадку вихідні дані відповідають даним, що знаходяться в базах даних і фази попередньої обробки та перетворення включають вилучення цих даних з бази даних і перетворення їх в структурований формат, який використовується програмним забезпеченням візуалізації.

Візуальне відображення. Ключові проблеми цього процесу полягають у визначенні того, які візуальні структури будуть використані для відображення даних і їх розташування в області відображення. Як вже було згадано, абстрактні дані не обов'язково мають реальне місце розташування у фізичному просторі. Існує декілька

типів абстрактних даних, які за своєю природою можуть легко знайти просторове розташування. Наприклад, дані, взяті зі станції моніторингу забруднення атмосфери можуть легко знайти положення на географічній карті, враховуючи, що станції моніторингу, які здійснюють вимірювання знаходяться в конкретному пункті на цій території. Те ж саме можна сказати і про дані, що отримано від топологічних структур, таких як трафік даних у комп'ютерній мережі. Однак існує кілька типів даних, які належать сутностям, які не мають природного географічного або топологічного позиціонування. Наприклад, це можуть бути бібліографічні посилання в наукових текстах, рівень споживання автомобільного палива, або зарплата різних фахівців в компанії. Такі типи даних не мають відповідного фізичного простору.

Тому необхідно визначити візуальні структури, які відповідають тим даним, які треба представити візуально. Цей процес називається візуальним відображенням. Треба визначити три структури [3, 5]:

- просторова підкладка;
- графічні елементи;
- графічні властивості.

Просторова підкладка визначає розміри у фізичному просторі, де створюється візуальне відображення. Просторова підкладка може бути визначена в вираженнях осей. В декартовому просторі, просторова підкладка відповідає осям x і y . Кожна вісь може мати різні типи, залежно від типу даних, які на ній відображаються. Зокрема, осі можуть бути кількісними (коли є метрика, асоційована зі значенням, що викладається на осі); порядковими (коли значення осі асоційовані з порядком, який відповідає порядку даних).

Графічні елементи являють все видиме, що з'являється в просторі. Існує чотири можливих типи візуальних елементів: точки, лінії, поверхні і об'єми (див. рис. 2).

Точки	Лінії	Поверхні	Об'єми
 	 	 	

Рис. 2. Приклади графічних елементів

Графічні властивості це властивості графічних елементів, до яких є чутливими сітківці ока. Найбільш поширеними графічними властивостями є розмір, орієнтація, колір, текстура і форма. Вони застосовуються для графічних елементів і визначають властивості візуального компонування (див. рис. 3).

Розмір	Орієнтація	Колір	Текстура	Форма
  	  	  	  	  

Рис. 3. Приклади графічних властивостей

З точки зору візуального сприйняття людини, не всі графічні властивості об'єктів поведуться однаково. Деякі графічні властивості є більш ефективними, ніж інші, з точки зору кількісних значень. Клівленд і Мак-Гілл [4] провели дослідження, щоб оцінити точність, з якою люди можуть сприймати кількісні значення, у якому зіставляються графічні елементи з різними властивостями і просторові підкладки. Дослідження визначило класифікацію, що представлено на рис. 4, з якої можна зробити висновок, що просторове позиціонування є одним з найбільш точних способів сприйняття кількісної інформації.

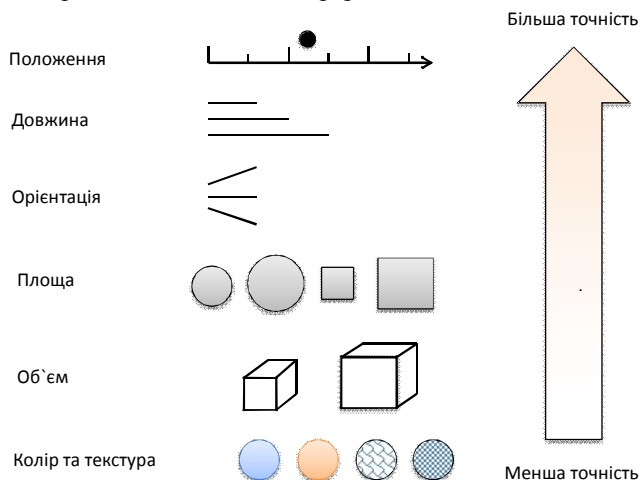


Рис. 4. Точність в сприйнятті кількісних значень для деяких графічних і просторових елементів

Зовнішній вигляд. Зовнішній вигляд є кінцевим результатом процесу генерації візуалізації інформаційних структур. Він є результатом відображення структури даних на візуальну структуру, створюючи візуальне уявлення в фізичному просторі за допомогою комп'ютера.

Зовнішній вигляд характеризується важкою проблемою - кількість даних, які треба представити є занадто великою для наявного простору. З цією проблемою, ми стикаємося досить часто, тому що реальні ситуації пов'язані з великими обсягами даних. У цих випадках, коли площа пристрою відображення занадто мала для відстеження всіх елементів візуального представлення, використовуються методи масштабування, панорамування, прокрутки, фокусування та контексту, віртуальні лінзи.

Основна проблема при розробці візуального представлення полягає у створенні візуального відображення, що, з одного боку, відтворює інформацію, що закодована в даних і, з іншого боку, полегшує здобуття заданої мети користувачем. Однак, немає ніякого способу дізнатися який тип візуального уявлення підходить для тих чи інших даних. Це залежить від характеру даних, того типу користувачів, для яких вони призначені, використання інформації та ін.

Процедура створення програмного забезпечення візуалізації абстрактних інформаційних структур повинна включати наступні етапи:

1. Визначення проблеми разом з потенційними користувачами візуального представлення. Визначення їх ефективної потреби. Це необхідно, щоб чітко визначити, які дані повинні бути представлені. Чому це представлення потрібно? Це необхідно для пошуку нової інформації? Або це потрібно, щоб довести гіпотези? Слід мати на увазі людний фактор, специфічний для цільової аудиторії, яка буде використовувати програмне забезпечення. Це буде впливати на вибір використовуваних візуальних моделей, щоб користувачі могли зрозуміти інформацію.

2. Вивчення характеру даних. Дані можуть бути кількісними (наприклад, список цілих або дійсних чисел), порядкові (дані без числової природи, але які мають свій власний внутрішній порядок, наприклад, дні тижня), або категоричний (дані, які не мають внутрішнього порядку, наприклад, імена людей або міст). Різні візуальні відображення можуть бути доречними, залежно від типу даних.

3. Кількість вимірів. Число вимірів даних, що повинні бути представлені, визначає тип представлення, яке треба використовувати. Атрибути можуть бути незалежними або залежними. За кількістю залежних атрибутів, існують дані, які називаються одновимірними (один вимір змінюється по відношенню до іншого), двовимірні (існує два залежні виміри), тривимірні (три залежні виміри), або багатовимірний (чотири і більше виміри, які змінюються відносно до незалежних).

4. Структури даних. Вони можуть бути лінійними (дані представляються в лінійні структури даних, такі як вектори, таблиці, колекції і т.д.), часовими (дані, які змінюються в часі), просторові або географічні (дані, які мають відповідність з чим-то фізичним - карта, план, тощо), ієрархічні (дані щодо сутностей, які організовані ієрархічно, наприклад, блок-схеми, структура файлів на диску і т.д.), і мережі (дані, які описують відносини між суб'єктами).

5. Тип взаємодії. Це визначає статичність візуального уявлення (наприклад, зображення, що надруковане на папері або зображення, представлене на екрані комп'ютера, але без змоги змін користувачем), його трансформованість (коли користувач може контролювати процес зміни і перетворення даних), або його маніпулятивність (користувач може управляти і змінювати деякі параметри, які регулюють генерацію зовнішнього виду). Модель, що представлено на рис. 1 ілюструє, на якому рівні обробки даних ці типи взаємодій можуть застосовуватись.

Елементи щойно описані, зведено в табл. 1.

Таблиця 1.

Змінні, які необхідно врахувати при розробці візуальних представлень

Проблема	Тип даних	Розмірність	Інформаційна структура	Тип взаємодії
Повідомлення	Кількісний	Одновимірний	Лінійна	Статична
Вивчення	Порядковий	Двовимірний	Часова	Трансформована
Підтвердження	Категоріальний	Тривимірний	Просторова	Маніпульована
		Багатовимірний	Ієрархічна	
			Мережева	

Кожен з можливих варіантів, що описані тут, може вказувати на використання конкретного методу візуалізації (див. рис. 5). Крім того, правильний вигляд повинен також визначити відповідні інструменти для оцінки впливу запланованого візуального подання на продуктивність роботи користувачів.

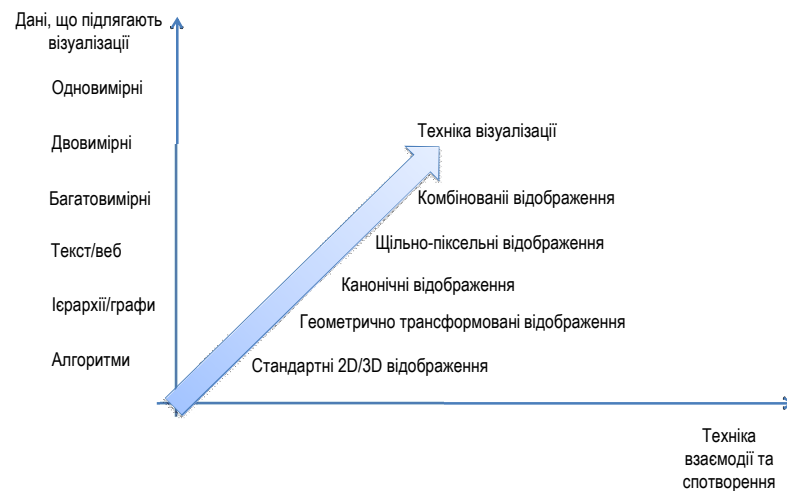


Рис. 5. Класифікація методів візуалізації інформаційних структур

Висновки

Дослідження великих наборів даних є важливою, але тяжкою проблемою. Методи візуалізації інформаційних структур можуть допомогти вирішити цю проблему. Візуальні дослідження даних мають високий потенціал і багато додатків, повинні використовувати технології візуалізації інформації для більш якісного аналізу даних. Майбутня робота буде включати в себе тісну інтеграцію візуалізації з традиційними методами таких дисциплін, як статистика, машинне навчання, дослідження операцій та моделювання. Інтеграція технологій візуалізації і цих методів зможуть поєднати алгоритми швидкого інтелектуального аналізу даних з інтуїтивними здатностями людського розуму, що дозволить поліпшити якість та швидкість процесу інтелектуального аналізу даних. Візуальні методи аналізу даних також повинні бути тісно інтегровані з системами, що використовуються для управління величезною кількістю спорідненої та слабоструктурованої інформації, включаючи управління базами даних і систем сховища даних.

Література

1. Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In Proceedings of the 1996 IEEE Symposium on Visual Languages, 1996.
2. А. Ю. Зинovieв Визуализация многомерных данных - Красноярск: Изд-во КГТУ, 2000. - 168с.
3. Card, S., Mackinlay, J., Shneiderman, B. Readings in Information Visualization: Using Vision to Think, Morgan Kaufmann, 2007.
4. Andrews, K., and Kasanicka, J. A comparative study of four hierarchy browsers using the hierarchical visualisation testing environment (HVTE). In Proceedings International Conference Information Visualization (IV) (2007), IEEE CS Press, pp. 81–86.
5. Riccardo Mazza. Introduction to Information Visualization. Springer-Verlag London Limited 2009, 149p.

Визуализация информации включает в себя визуальное представление абстрактных информационных пространств и структур для облегчения их быстрого понимания и применения. Информационные структуры включают иерархии, сети, многомерные таблицы и т.п. Методы, основанные на визуализации информационных структур находят все большее применение в различных областях науки.

Ключевые слова: визуализация, информация, информационная структура, граф, сеть, программное обеспечение, графическое представление

Information visualization includes a visual representation of abstract information spaces and structures to facilitate their quick understanding and application. Information structures include hierarchies, networks, multi-dimensional tables, etc. Methods based on the data structures visualization are widely used in all areas of science.

Key words: visualization, information, information structure, graph, network, software, graphical representation

Дядичев В.В. - докт.техн.наук., професор. завідувач кафедри комп'ютерних наук СХУ ім. В. Даля
Сафонов К.М. - старший викладач кафедри комп'ютерних наук СХУ ім. В.Даля

Рецензент: Погорелов О.О. докт.техн.наук., професор СХУ ім. В. Даля