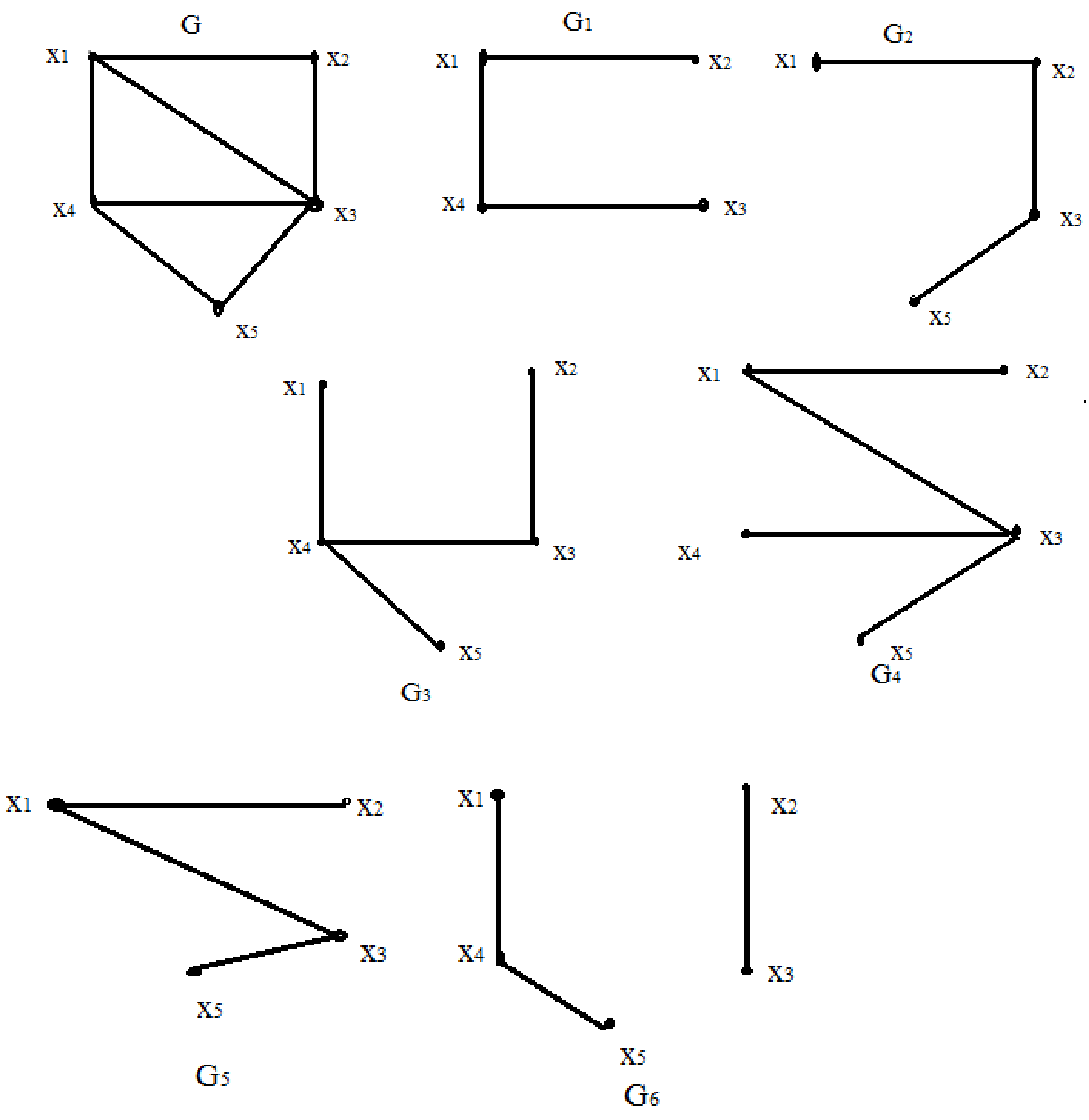
Дерева. Остови, кодерева.

Графи, які зустрічаються в багатьох прикладних задачах, є зв`язними. Серед зв`язних графів дерева, що мають просту структуру, є найважливішими.



Граф без циклів називають *<u>ациклічним</u>*, або *<u>лісом</u>*. *<u>Деревом</u>* називається зв`язний ациклічний граф. Таким чином, компонентами зв`язності лісу є дерева.

:Граф, у якого $m\text{-}1=n$ (n – кількість вершин, m- кількість ребер), називають *<u>деревоподібним.</u>*

Остов графа G – це дерево графа G, яке вміщує всі вершини графа G. Зв`язний підграф дерева T називають *піддеревом* дерева T.

Для графа G підграфи G_1 і G_2 є деревами, графи G_3 і G_4 – остовами графа G.

Кодерево T' остова T графа G є підграфом графа G , який вміщує всі вершини графа G і тільки ті ребра графа G, які не входять в T. На рисунку зображені кодерева G_5 і G_6. Кодерево може бути незв`язним. Ребра остова T називають *гілками*, а ребра відповідного кодерева $T'$ - *хордами* або зв`язками.

Якщо в ациклічному графі x_i та x_j несуміжні вершини, тобто $l_i = (x_{i,} x_j) \notin E$, а граф G+1 має лише один простий цикл, то його називають *субциклічним* графом.

Властивості дерева.

Наступна теорема встановлює, що дві з чотирьох властивостей: зв`язність, дерево подібність, ациклічність, субциклічність, характеризують граф, як дерево.

Теорема. Нехай граф G=(V,E) – граф з n вершинами, m ребрами, k компонентами зв`язності та z простими циклами. Нехай l –ребро, що зв`язує довільну пару несуміжних вершин в G. Тоді наступні твердження еквівалентні:

1. G-дерево, тобто зв`язний граф без циклів; K(G)=1; Z(G)=0.
2. Довільні дві вершини, з`єднані в G єдиним простим ланцюгом.
3. G-зв`язний граф, довільне ребро є мостом; K(G)=1 та існує ребро l таке, що K(G-1)>1/
4. G- зв`язний граф і деревоподібнийK(G)=1;m=n-1.
5. G -ациклічний та деревоподібний Z(G)=0;m=n-1.
6. G –ациклічний та субциклічний Z(G)=0; Z(G+1)=1.
7. G- зв`язний субциклічний та неповний; K(G)=1; $G \neq K_n, n \geq 3$; Z(G+*l*)=1.
8. G-деревоподібний субцикічний та m=n-1, $G \neq K_1 \cup K_3, G \neq K_{2 \cup K_3}$, Z(G+*l*)=1.

   Доведнмо, наприклад, що з 5 слідує 1. Граф без циклів, тобто його компоненти – дерева. Нехай їх k. Тоді:

   $m = \sum_{i=1}^{k} m_i = \sum_{i=1}^{k} (n_i - 1) = \sum_{i=1}^{k} n_i - k = n - k$, але m=n-1, тобто k=1.

   Наслідок. Довільне нетривіальне дерево має хоча б дві висячі вершини.

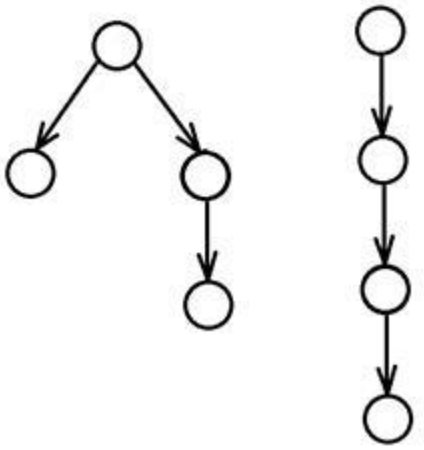
Орієнтовані дерева.

Орієнтоване дерево та ієрархія – це рівнозначні поняття. Дерева є абстракцією ієрархічних відношень.

Орієнтованим деревом або *ордеревом*, або *кореневим деревом* називають граф, що має такі властивості:

1. Існує єдина вершина x_0 графа з півстепенню заходу $d_+(x_0) = 0$, яку називають коренем дерева.
2. Кожна інша вершина має пів степінь заходу $d_+(x_i) = 1, i \neq 0$.
3. Кожна вершина x_i досягається з кореня, тобто існує орієнтований шлях від кореня x_0 до всіх інших вершин графа.

На рисунку зображені різні ордерева.

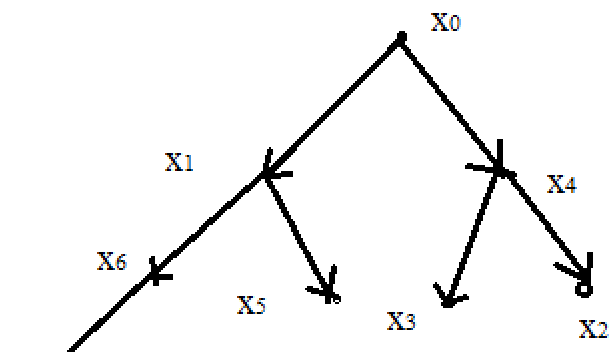


Теорема. Орієнтоване дерево має такі властивості:

1. m=n-1.
2. Якщо в ордерові відмінити орієнтацію, орграф стає вільним деревом.
3. В ордерові відсутні контури.
4. Для кожної вершини існує єдиний шлях, що веде до неї з кореня.
5. Підграф, що визначається множиною вершин, що досягаються з вершини x_i є ордеревом з коренем x_i. Таке ордерево називають під деревом вершини x_i.
6. Якщо у вільному графі довільну вершину призначити коренем, то отримаємо ордерево.

Кінцева вершина ордерева називається *листом*, шлях з кореня до листа називають гілкою. Довжина найбільшої гілки ордерева називається його *висотою*. *Рівень точки* ордерева – довжина від кореня до цієї точки. Сам корінь має рівень 0.

Точки одного рівня утворюють *ярус* дерева. На рисунку висота дерева дорівнює 3, ярус рівня 1 утворюють вершини $\{x_4; x_1\}$, ярус рівня 2 утворюють вершини $\{x_6; x_5; x_3; x_2\}$.



Досить часто для ордерів використовується генеалогічна термінологія.

Точки, які можна досягнути з точки x_i називають нащадками вузла x_i. Нащадки утворюють піддерево. Якщо існує в дереві дуга (x_i, x_j) , то x_i називають батьком точки x_j, а x_j сином точки x_i. Синів однієї точки називають братами.

Дамо ще одне означення ордерева.

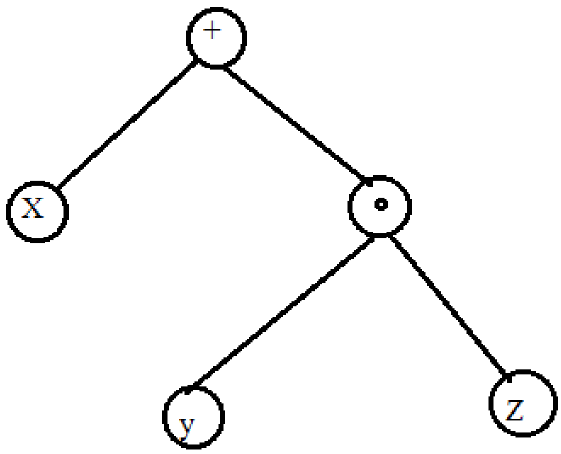
Ордерево T – це скінчена множина вершин таких, що є лише одна така точка x_0 - корінь даного дерева, а всі інші разом з коренем вміщуються в r-множинах $T_1, T_2 ..., T_r$, що попарно не перетинаються, кожна з яких утворює ордерево $(r \geq 0)$. $T = \{\{x_o\}, T_1, T_2, \ldots, T_r\}$.

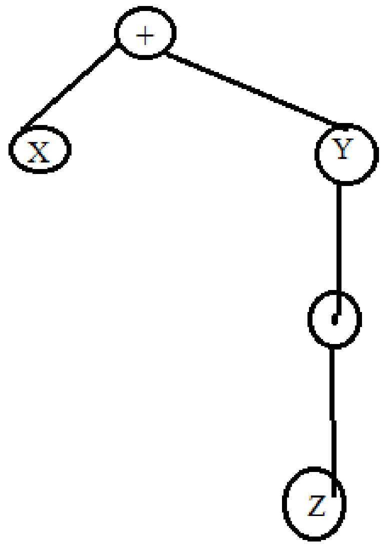
$T_i\left(i = \overline{1, r}\right)$ - піддерева. Якщо їх порядок зафіксовано, ордерево називають впорядкованим.

У програмуванні інтенсивно використовують ордерева для запису алгебраїчних виразів, побудови алгоритмів, а структура вкладень каталогів та файлів у учасних операційних системах є впорядкованим деревом. Це відображається навіть в терміні – «кореневий каталог дисків».

На практиці для зображення дерев прийнята домовленість про те, що корінь дерева знаходиться зверху, а всі стрілки дуг направлені вниз, тому стрілки можна не зображати.

Зобразимо через дерева, наприклад, алгебраїчний вираз $x + yz$ та $(x + y)z$



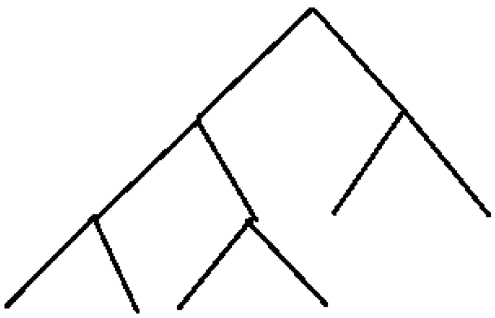


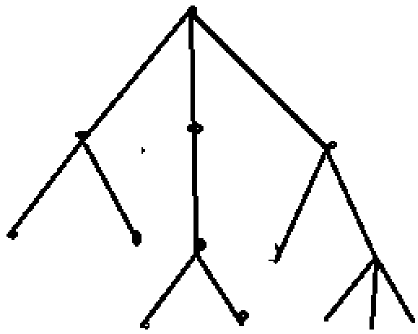
Дерева з мінімальною довжиною зважених шляхів.

Деякі визначення:

m –деревом називається ордерево, в якому напівступінь виходу кожної вершини не перевищує m.

На рисунку зображене двійкове, або 2-арне, або бінарне дерево.





На другому рисунку 3-дерево або тернарне.

Найбільше застосування в обчислювальній техніці при аналізі алгоритмів в методах пошуку інформації, організації індексування даних і таке інше мають бінарні або двійкові дерева.

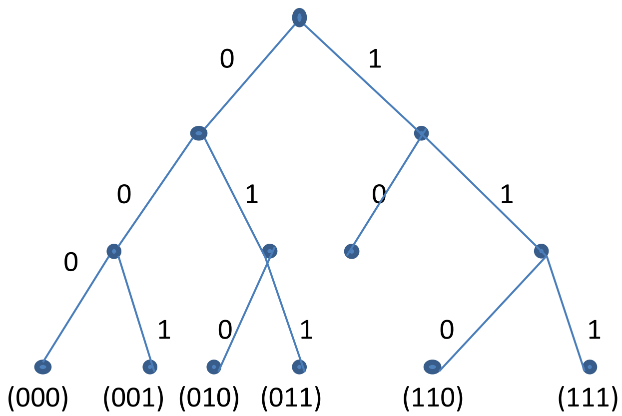
Однією з таких задач є наступна:

Дано n-ваг $w_1, w_2, \ldots, w_n$ та n довжин $l_1, l_2, \ldots, l_n$ листами $x_1, x_2, \ldots, x_n$ таке, що:

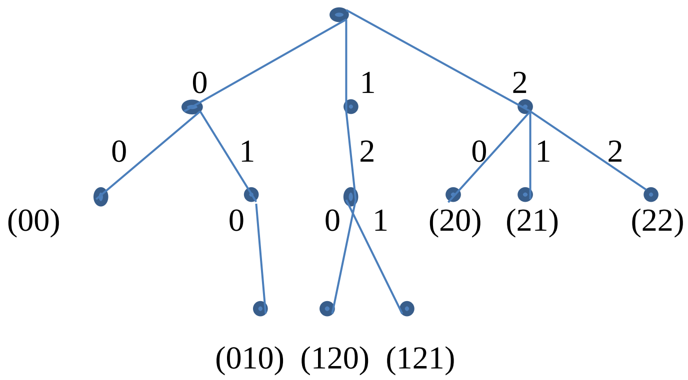
1. $w_i (1 \le i \le n)$ - вага вершини x_i;
2. $l_i (1 \le i \le n)$- довжина шляху з кореня дерева в x_i;
3. $\sum_{i=1}^{n} w_i l_i$, якою називають довжиною зважених шляхів – мінімальна.

Нехай T m-дерево, $M = \{0,1,2,\ldots, m-1\}$ - алфавіт, що складається з m букв. Нехай ми приписуємо кожному ребру дерева T букву з алфавіту M так, що не може бути двох ребер, що виходять з однієї й тієї самої вершини, помічених однією буквою. Тоді ми можемо приписати кожній вершині дерева T слово, що утворюється конкатенацією (ланцюгом) букв, якими помічені ребра, що зустрічаються при руху з кореня в дану вершину. Такі слова, прописані листами дерева T, називають кодовими словами, і кажуть, що вони утворюють префіксний код.

Так бінарному дереву (рис.) відповідає префікс ний код $\{000, 001, 010, 011, 110, 111\}$.



.



Для тернарного дерева кодовими словами є $\{00, 010, 120, 121, 20, 21, 22\}$.

Нехай ми закодували L одиниць інформації $M_1, M_2, ..., M_L$ словами префіксного коду. Передаючи її по каналах зв`язку, отримаємо конкатенацію кодових слів що відповідають переданій інформації.

Щоб із цієї послідовності виділити інформацію, необхідно розкласти її на слова префіксного коду. Процес розкладання послідовності на кодові слова називають декодуванням, і його можна вирішити за допомогою дерева, що відповідає префіксному коду.

Наприклад, розглянемо послідовність 120202200, яка утворена конкатенацією деяких слів префіксного коду тернарного дерева.

Щоб декодувати цю послідовність, розглянемо її букви зліва направо. Відповідно з цим, ми рухаємося по дереву, починаючи з кореня, вздовж ребер, які відповідають розглянутим буквам, поки не дійдемо до листа дерева. Кодове слово, що відповідає цьому листу, є першим словом в даній послідовності. Ми отримаємо 120.

Далі повторюємо цю процедуру для 202200 і виділяємо слова 20, 22, 00 друге, третє, четверте кодові слова послідовності 120202200.

З процесу декодування зрозуміло, що вартість декодування кодового слова пропорційна кількості букв у слові.

Якщо W_i - частота появи інформації M_i, то вартість декодування залежить від суми $\sum_{i=1}^{n} W_i l_i$; де l_i - довжина шляху від кореня до листа, що відповідає інформації M_i. $\sum_{i=1}^{n} W_i l_i$ - довжина зважених шляхів дерева. Тому розв`язують задачу побудови m-дерева з мінімальною довжиною зважених шляхів на заданій множині ваг.