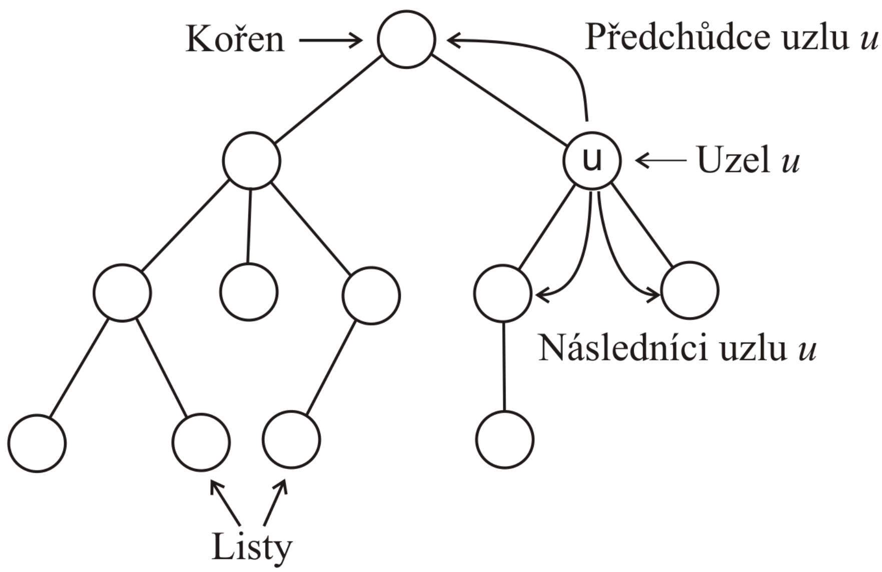
Stromy

Z nelineárních datových struktur jsou v algoritmech nejpoužívanější stromy. Stromy jsou specifickým případem matematických grafů. Terminologií grafů bychom je popsali jako obyčejné acyklické grafy. Skládají se z uzlů a hran. Při použití stromů v algoritmech jsou v uzlech uloženy datové prvky, nad kterými algoritmus probíhá. Hrany reprezentují vztahy mezi uzly (prvky), které jsou základem daného algoritmu.

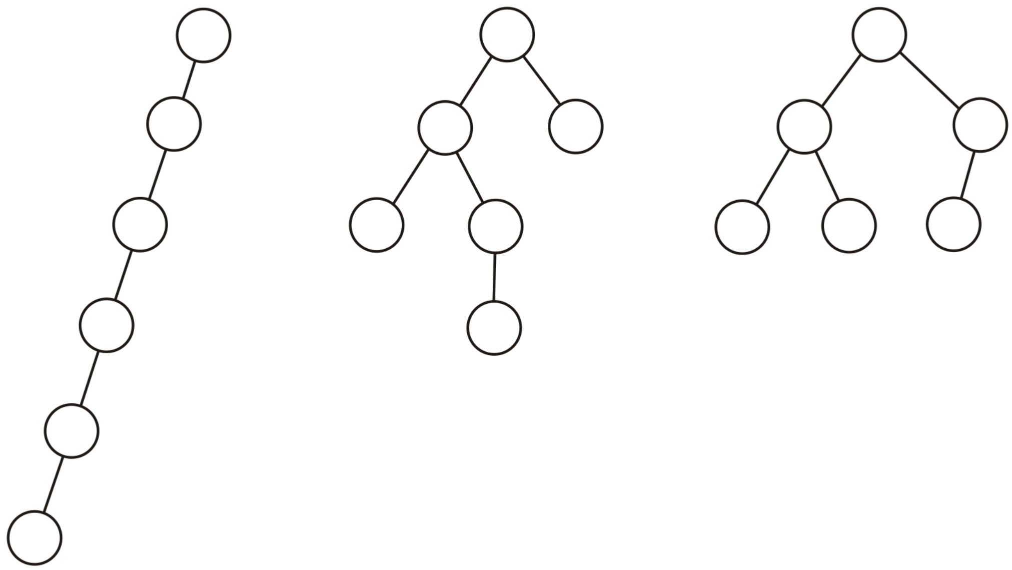
Strom se kreslí směrem shora-dolů (někdy také zleva-doprava, je-li příliš široký). Zcela nahoře je první uzel stromu, který se nazývá kořen. Pod ním jsou uzly, které jsou jeho následníci. Jsou s ním spojeny hranami. Tyto uzly mohou mít rovněž následníky, které jsou opět pod nimi a jsou s nimi opět spojeny hranami. Uzly, které nemají žádné následníky, nazýváme listy. Každý uzel vyjma kořene je hranou spojen s právě jedním uzlem, který je nad ním. Tento uzel je jeho předchůdce.

Na kreslení uzlů a hran nejsou žádná zvláštní omezení. Uzly většinou kreslíme jako kružnice, hrany jako rovné čáry.



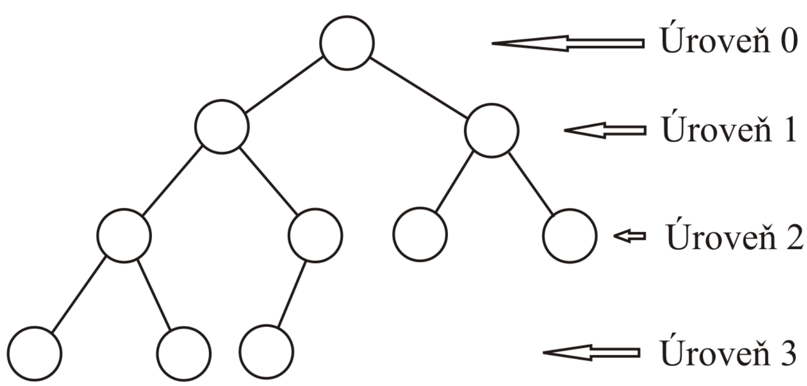
Běžně používané stromy mají zpravidla omezeno, kolik může mít uzel následníků. V tomto ohledu nejjednodušší stromy jsou stromy, v nichž každý uzel může mít nejvýše dva následníky. Používají se poměrně často a říkáme jim binární stromy.

Při použití stromů v algoritmech si uchováváme ukazatel na kořenový uzel. K libovolnému uzlu se pak dostaneme tak, že začneme od kořene a postupně po hranách přecházíme k nižším uzlů, až dojdeme k žádanému uzlu. Přechod od nějakého uzlu po hraně k jeho následníkovi považujeme za jednu základní operaci. Počet operací potřebný k tomu, abychom se od kořene dostali k danému uzlu, je roven počtu hran, které jsou na cestě od kořene k tomuto uzlu. Tento počet nazýváme vzdáleností uzlu od kořene. Maximum ze vzdáleností uzlů od kořene stromu nazveme výškou stromu. Výška stromu je tedy vzdálenost kořene od listů, které jsou ve stromu nejníže. Zřejmě čím má strom při daném počtu uzlů menší výšku, tím je to výhodnější, neboť tím nižší je maximální délka cesty od kořene k uzlům stromu. Na následujícím obrázku jsou tři binární stromy. Všechny mají stejný počet uzlů – šest.



Levý strom je nejméně výhodný. V podstatě je to seznam a o seznamu víme, že časová složitost přístupu k jeho uzlům je lineární. Naproti tomu strom zcela vpravo má tu vlastnost, že při daném počtu uzlů má nejmenší možnou výšku. Je to vyvážený binární strom.

Vyvážený binární strom je takový binární strom, který má ve všech úrovních maximální možný počet uzlů vyjma poslední úrovně, která může být zaplněna jen zčásti. Úrovní přitom tady rozumíme všechny uzly, které mají stejnou vzdálenost od kořene, tedy mají stejnou vodorovnou úroveň při běžném nakreslení stromu shora-dolů. Tuto vzdálenost nazveme číslem úrovně. Podíváme-li se na následující příklad vyváženého binárního stromu



můžeme z něho odvodit, jaké počty uzlů budou obecně v jednotlivých úrovních vyváženého binárního stromu výšky *h*.

Číslo úrovně	Počet uzlů v ní
0	1
1	2
2	2^2
…	…
h-1	2^{h-1}
h	1 .. 2^h

Odtud pro počet uzlů *n* v celém stromu dostáváme vztah:

$$1+2+2^2+\ldots+2^{h-1}+1 \le n \le 1+2+2^2+\ldots+2^{h-1}+2^h \quad .$$

Jeho postupnými úpravami (použitím součtu geometrické řady):

$$2^h \le n \le 2^{h+1}-1$$

$$\frac{n+1}{2} \le 2^h \le n$$

$$\log_2\left(\frac{n+1}{2}\right) \le h \le \log_2(n)$$

$$\log_2(n+1)-1 \le h \le \log_2(n)$$

Z tohoto plyne důležitý závěr, že ve vyváženém binárním stromu výška stromu logaritmicky závisí na počtu uzlů v něm, což můžeme vyjádřit jako složitost Θ(log(n)).