

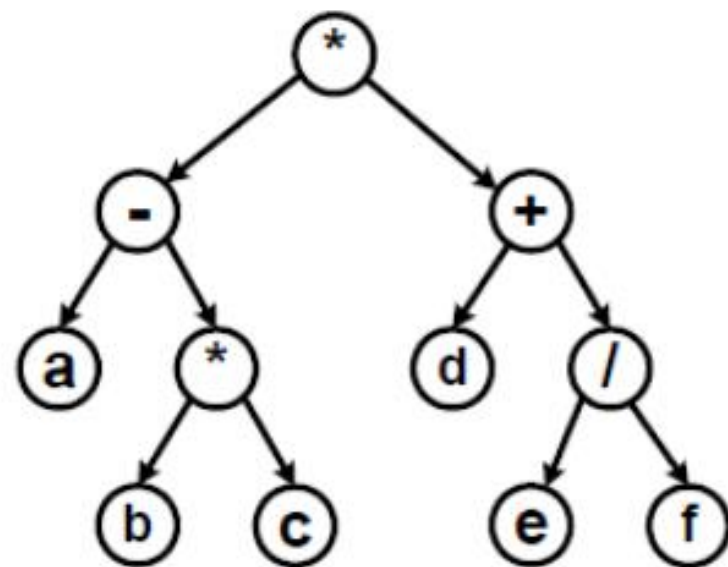
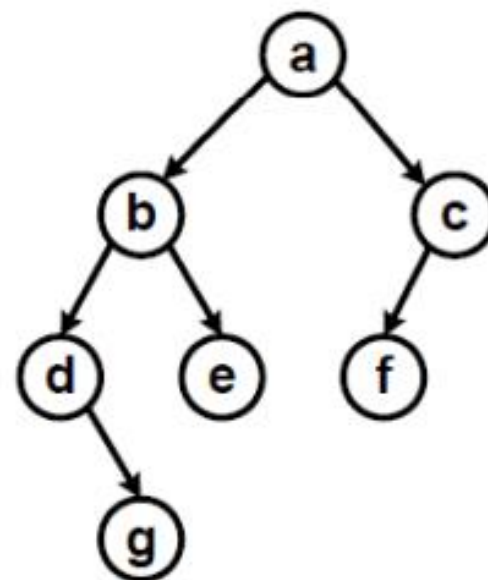
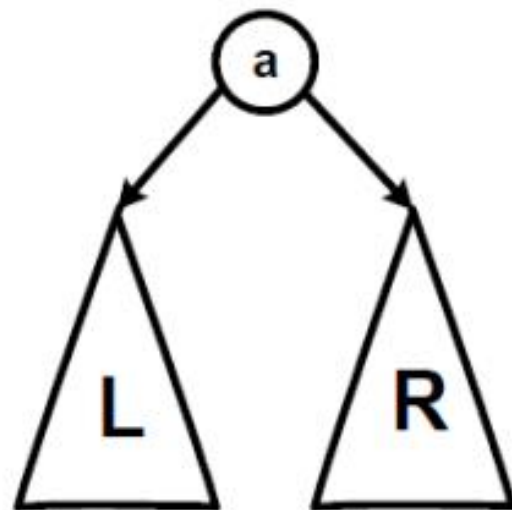
DUOMENŲ STRUKTŪROS IR ALGORITMAI

MARIUS GŽEGOŽEVSKIS

HERARCHINĖS DUOMENŲ STRUKTŪROS MEDŽIAI (angl. TREES)

Medžiai – tai netiesinės duomenų struktūros, kurių pagrindinis privalumas labai greita paieška.

- ✓ Medis – tai ciklą neturintis orientuotas grafas.
- ✓ Medį sudaro baigtinis elementų skaičius, kurie vadinami viršūnėmis.
- ✓ Medžio viršūnė, į kurią neįeina jokia kita viršūnė, vadinama šaknimi.
- ✓ Į kiekvieną viršūnę, išskyrus šaknį, įeina ne daugiau nei viena viršūnė.
- ✓ Medžio dalis, išeinanti iš bet kurios viršūnės, vadinama šaka.



$$(a - b * c) * (d + e / f)$$

MEDŽIAI (angl. trees)

Iki šiol nagrinėtos duomenų struktūros buvo vienmatės - vienas duomuo sekė po kito. Tokias DS gan paprasta programuoti ir naudoti, tačiau taikymams jos gali būti neefektyvios, t.y. algoritmas, veikiantis jų pagrindu, gali būti per lėtas.

MEDŽIAI (angl. trees)

Todėl verta nagrinėti sudėtingesnes - dvimates, trimates ir pan. duomenų struktūras. Viena iš tokių yra medžio duomenų struktūra. Medis - tai duomenų aibė, kuriai apibrėžtos duomenų incidentumo ir eiliškumo operacijos:

MEDŽIAI (angl. trees)

- ✓ medžio (žymime T) elementai yra dvi aibės: viršūnių aibė V ir briaunų aibė E ;
- ✓ tarp V ir E apibrėžtas incidentumo santykis: viena briauna (elementas iš E) atitinka dvi ir tik dvi viršūnes (elementų iš V porą, elementų tvarka poroje yra fiksuota);
- ✓ viena viršūnė medyje yra išskiriama ir vadinama šaknimi;
- ✓ bet kurias dvi viršūnes jungia vienas ir tik vienas kelias (kelias yra briaunų seka, kurioje kiekvienos dvi gretimos briaunos turi bendrą viršūnę).

MANIPULIAVIMO SU DUOMENIMIS OPERACIJOS TOKIOS:

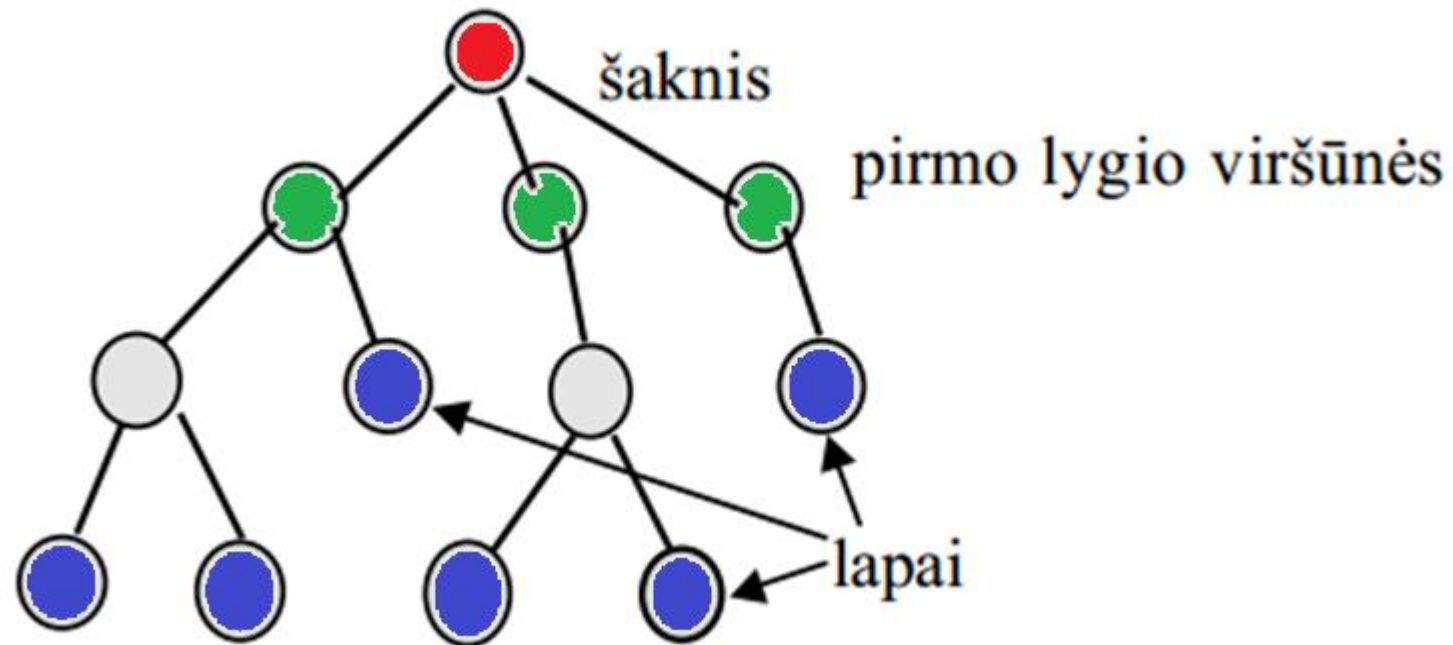
- ✓ inicializuoti medį (išskirti vietą būsimai medžio struktūrai kompiuterio atmintyje);
- ✓ įterpti elementą į medį (įterpti naują viršūnę);
- ✓ išmesti elementą iš medžio (pašalinti viršūnę);
- ✓ sujungti du medžius;
- ✓ išskaidyti medį į du medžius;
- ✓ sutvarkyti medį (išdėstyti medžio viršūnes taip, kad pagal tam tikrą viršūnių perrinkimo tvarką, jų reikšmės atitiktų norimas sąlygas);
- ✓ panaikinti medį (išlaisvinti medžiui skirtą kompiuterio atmintyje vietą).

MEDŽIAI (angl. trees)

Viršūnė medyje dažniausiai yra struktūrizuota (atitinka formuojamas algoritme reikšmes ar sąlygas, dažnai gan sudėtingas). Ji gali turėti vardą ir prie jos papildomą informaciją, taip pat tipą ar netgi duomenų struktūrą. Briauna irgi gali identifikuoti papildomą informaciją (pvz. svorį), gali turėti tipą. Kelias iš šaknies į kurią nors viršūnę yra vadinamas šaka.

MEDŽIO PAVYZDYS

Kelias iš šaknies į kurią nors viršūnę yra vadinamas šaka. Medis yra braižomas šakomis žemyn, šaknį dėstant viršuje.



MEDŽIAI (angl. trees)

Nors briaunos paprastai neturi krypties (medis yra neorientuotas grafas), tačiau programuojant briaunos realizuojamos rodyklėmis ir intuityviai joms priskiriama kryptis. Jei ši kryptis iš viršūnės rodo į kitą viršūnę žemyn nuo šaknies, tai sakoma, kad viršūnė turi sūnų.

N-ARINIS MEDIS

Briaunos kryptis į priešingą pusę (viršų) nurodo viršūnei jos tėvą. Viršūnė be sūnų vadinama lapu arba išorine viršūne. Visos kitos viršūnės vadinamos vidinėmis. Vieno tėvo sūnūs vadinami broliais (siblings). Jei kiekvienas tėvas turi ne daugiau kaip fiksuotą skaičių sūnų (tarkime, n), jis vadinamas n -ariniu medžiu.

DVEJATAINIAI MEDŽIAI

Algoritmų teorijoje ypač svarbūs yra **dvejetainiai medžiai**. Medis, prasidedantis kurioje nors viršūnėje, vadinamas pomedžiu. Miškas yra dviejų ir daugiau medžių aibė. Medis turi lygius: šaknis yra nulinio lygio, jos sūnūs yra pirmo lygio, šių sūnų sūnūs yra antro lygio ir t.t. Medžio aukščiu vadinamas lygių maksimumas (arba maksimalios šakos ilgis).

MEDŽIO SAVYBĖS

Jei n_1, n_2, \dots, n_k yra medžio viršūnės, ir n_i yra viršūnės n_{i+1} tėvas, tada kelias nuo viršūnės n_1 iki n_k bus vadinamas $k - 1$ keliu.

Iš šaknies į kiekvieną viršūnę veda vienintelis kelias.

Viršūnės, nutolusios nuo šaknies atstumu k , vadinamos k -ojo lygmens viršūnėmis.

MEDŽIO SAVYBĖS

Viršūnės, kurios neturi dukterinių viršūnių, vadinamos lapais.

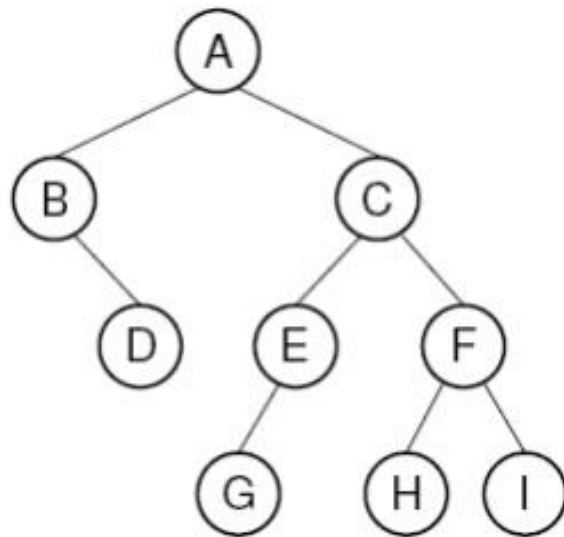
Medžio aukštis yra vienetu didesnis už gylį iki giliausios viršūnės.

Šaknies gylis lygus 0, šaknis yra nuliniam lygyje.

Medžiai gali būti: dvejetainiai, trejetainiai, ketvirtainiai ir t.t.

DVEJETAINIS MEDIS

Dvejetainiu vadinamas orientuotas medis, kuriame į kiekvieną viršūnę, išskyrus šaknį, įeina viena briauna, o išeina ne daugiau kaip dvi. Dvejetainis medis turi ne daugiau nei du viršūnes-vaikus



Medžio aukštis - 4

Viršūnės H gylis - 3

Šaknis - A

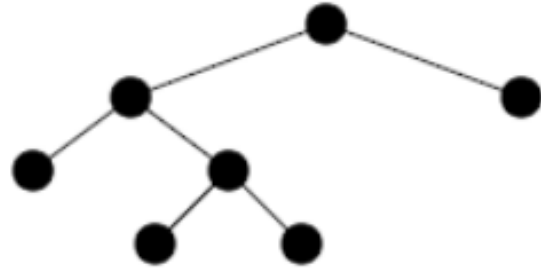
Lapai – D G H I

Pilnas ir subalansuotas medis

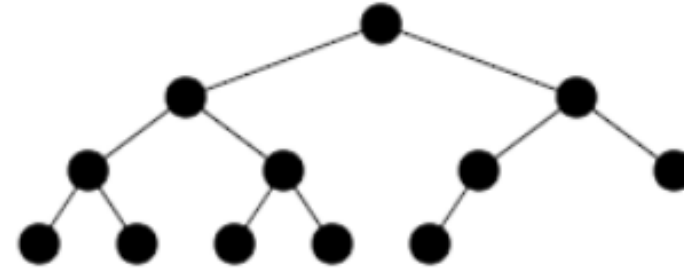
Pilnu (full) medžiu vadiname tokį medį, kuriame kiekviena viršūnė turi du vaikus arba yra lapas.

Subalansuotas medis – tai toks medis, kuriame viršūnės pildomos iš kairės į dešinę tol, kol užpildomas visas lygis ir tik tuomet pereinama į sekantį lygį.

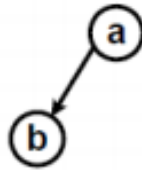
Subalansuoto medžio visi lygiai išskyrus paskutinį yra pilnai užpildyti. Paskutinis lygis pildomas iš kairės į dešinę.



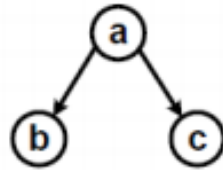
Pilnas (nesubalansuotas) medis.



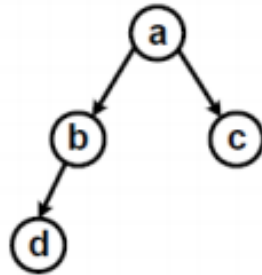
Subbalansuotas (nepilnas) medis.



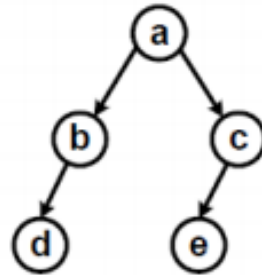
a)



b)



c)



d)

Subbalansuoti medžiai

Panagrinėkime algoritmą, dvejetainio medžio formavimui. Tarkime iš duomenų n_1, n_2, \dots, n_k reikia sudaryti dvejetainio medžio struktūrą.

Dvejetainio medžio formavimo algoritmas (1):

1. n_1 – įrašomas šaknyje.
2. n_2 lyginamas su n_1 . Jei $n_2 < n_1$, tuomet jungiamas prie kairės šakos, priešingu atveju – prie dešinės šakos.
3. Antras žingsnis kartojamas su kitais n_i elementais, pradedant lyginimą nuo šaknies ir tęsiant tol, kol elementas tampa lapu.

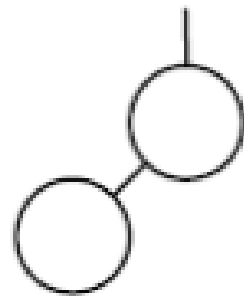
Dvejetainio medžio formavimas (2)

Mažiausio aukščio dvejetainio medžio formavimo algoritmas (2).
Viršūnės išdėstomos taip, kad būtų užpildytas maksimaliai visas žemiausias lygis pradedant pildyti vienodai nuo kairės kiekvienos viršūnės šakos, po to pereinant prie dešinėsios šakos.

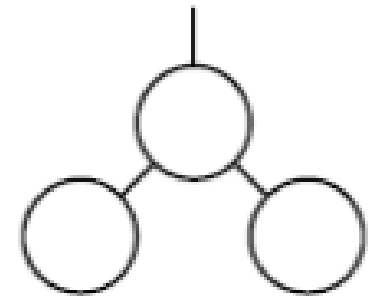
$n=1$



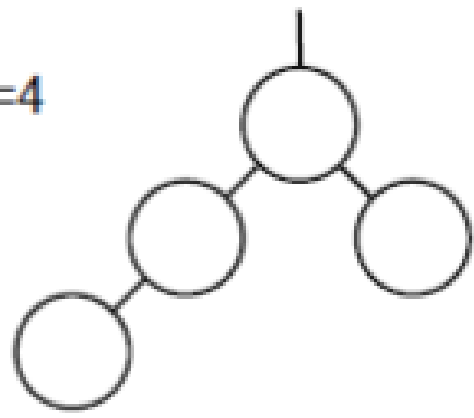
$n=2$



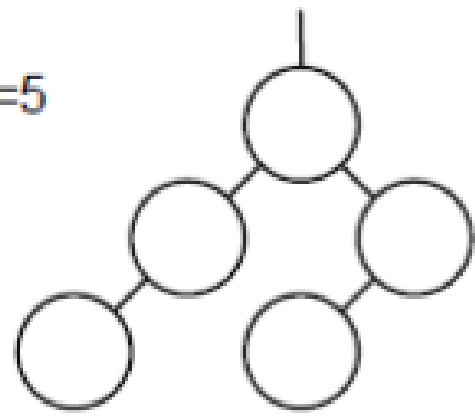
$n=3$



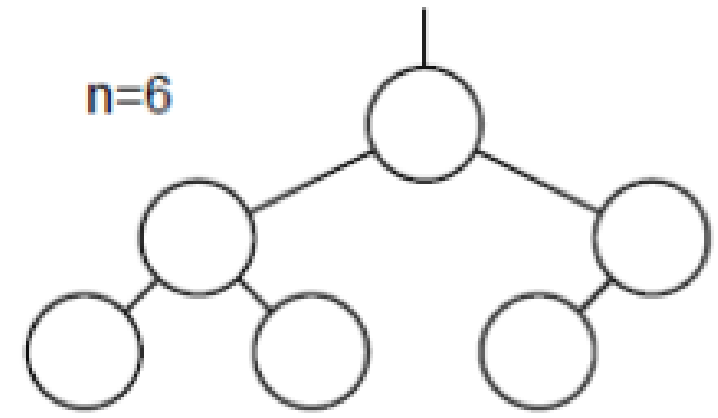
$n=4$



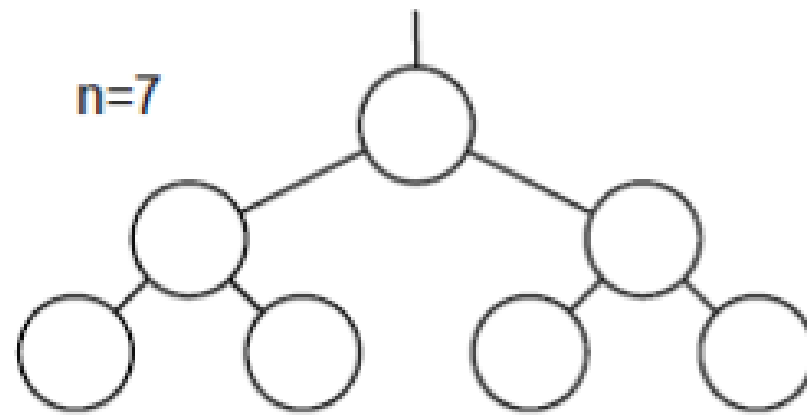
$n=5$



$n=6$



$n=7$



Pilnas subalansuotas medis

Kiek viršūnių N turi pilnas subalansuotas medis, kurio gylis h ?

$$N = 1 + 2^1 + 2^2 + \dots + 2^h \text{ Arba } N = 2^{h+1} - 1 \text{ kur } h = (\log_2 N)$$

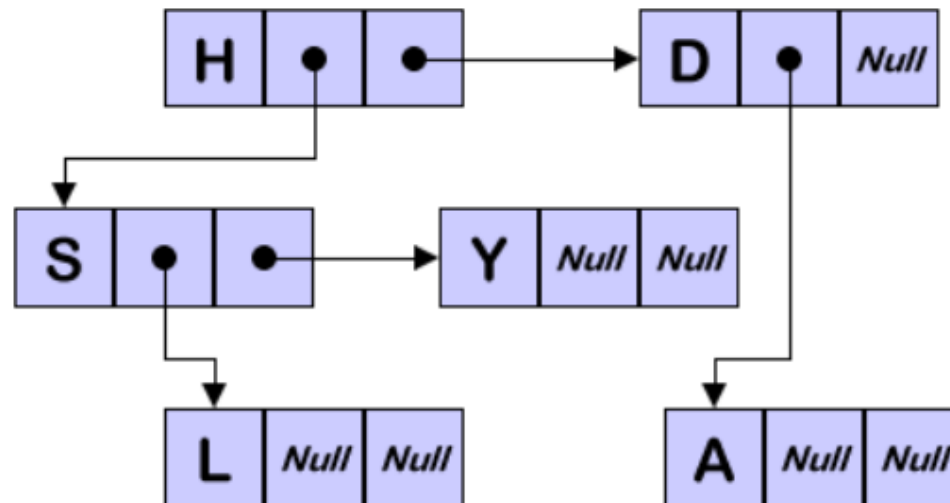
Norint surasti bet kurią viršūnę tokiaame dvejetainiame medyje, blogiausiu atveju tai užims $h+1$ arba $(\log_2 N)$ žingsnių.

DVEJETAINIO MEDŽIO ELEMENTAS

Dvejetaio medžio viršūnės elementas susideda iš duomenų ir dviejų rodyklių.

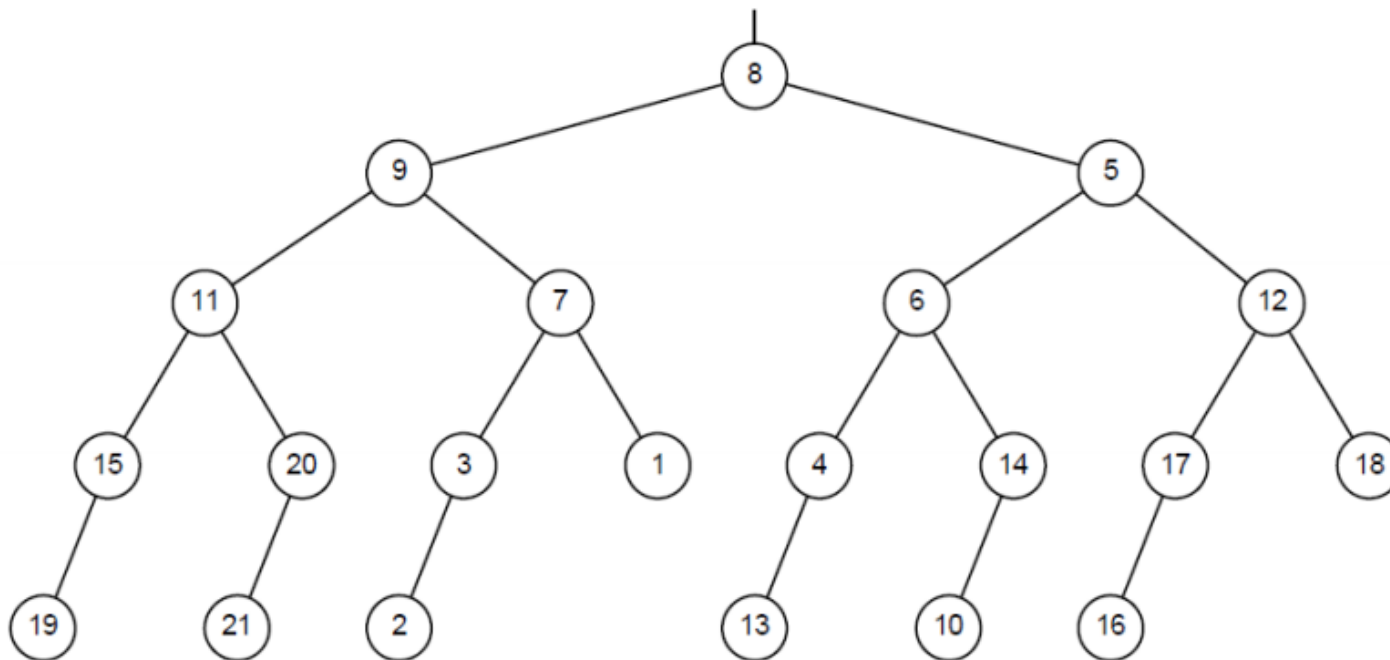


Dvejetaio
medis



DVEJETAINIO MEDŽIO PAVYZDYS

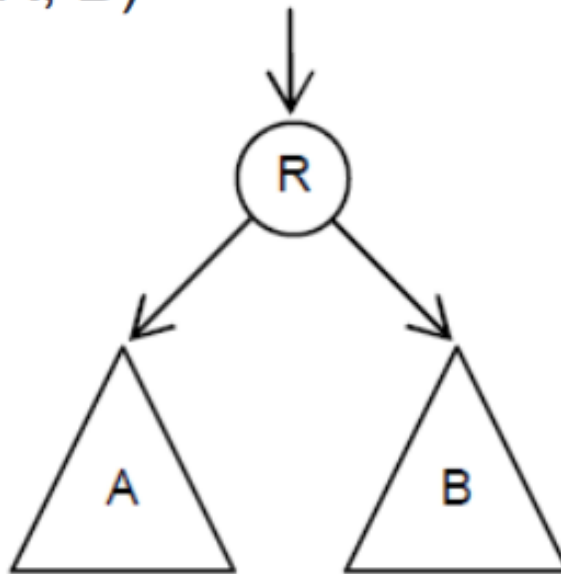
Iš 21 skaičiaus aibės suformuojamas dvejetainis medis :
(8 9 11 15 19 20 21 7 3 2 1 5 6 4 13 14 10 12 17 16 18)

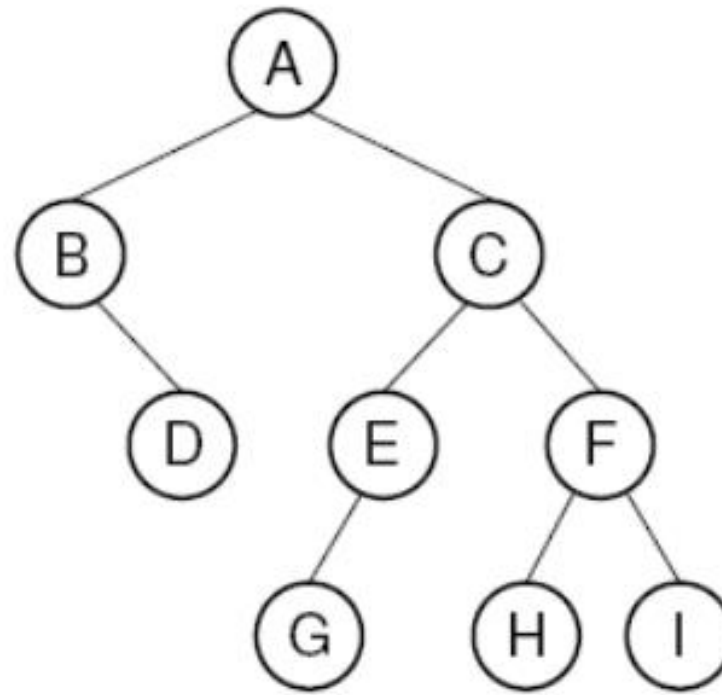


MEDŽIO APĖJIMO BŪDAI

Nuskaitant medžio struktūros viršūnės, reikia tai padaryti tik po vieną kartą. Viršūnės ir dviejų jos vaikų apėjimo tvarka gali _____ būti:

- Tiesioginė (R, A, B)
- Atvirkštinė (A, B, R)
- Vidinė (A, R, B)





Tiesioginis apėjimas:

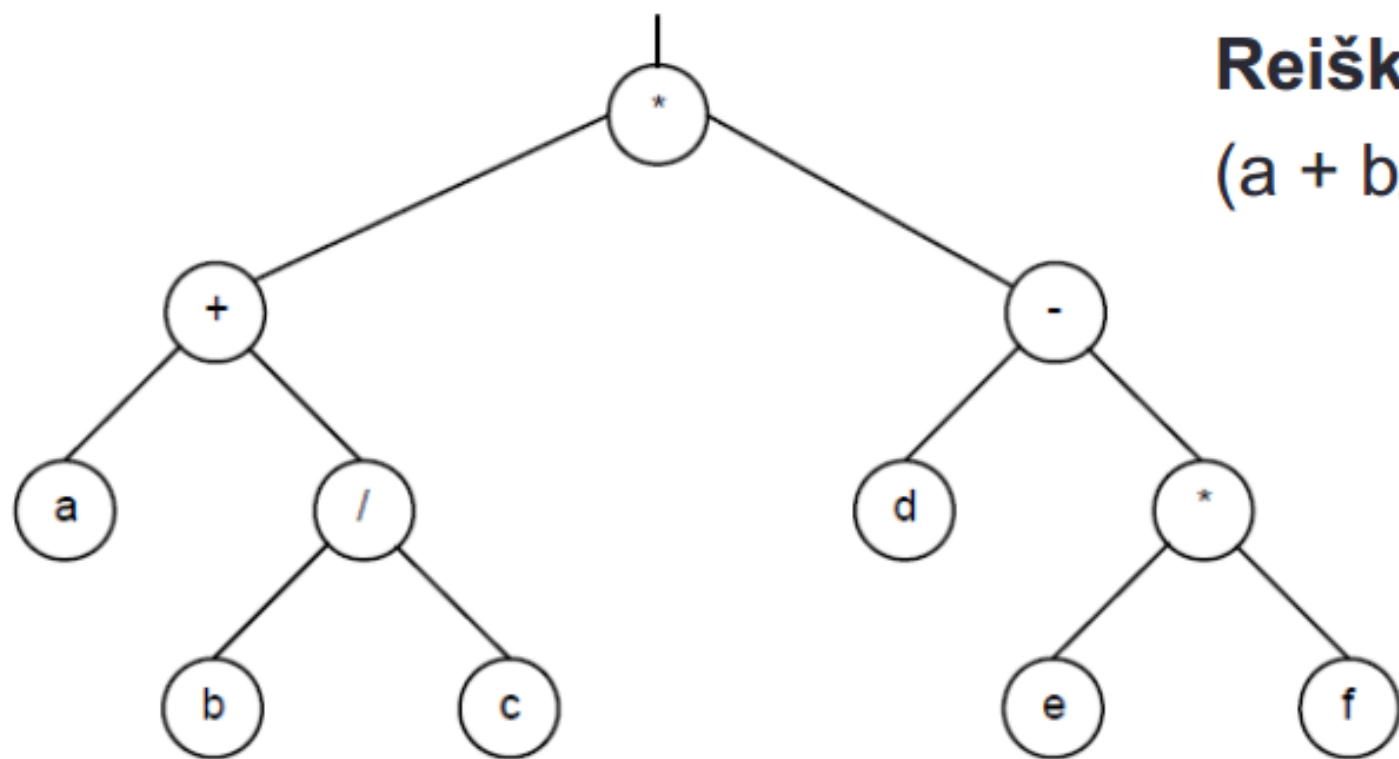
A B D C E G F H I

Atvirkštinis apėjimas:

D B G E H I F C A

Vidinis apėjimas:

D B A G E C H F I



Reiškinys:

$(a + b/c) * (d - e*f)$

1. Tiesioginis:

*** + a / b c - d * e f**

2. Vidinis:

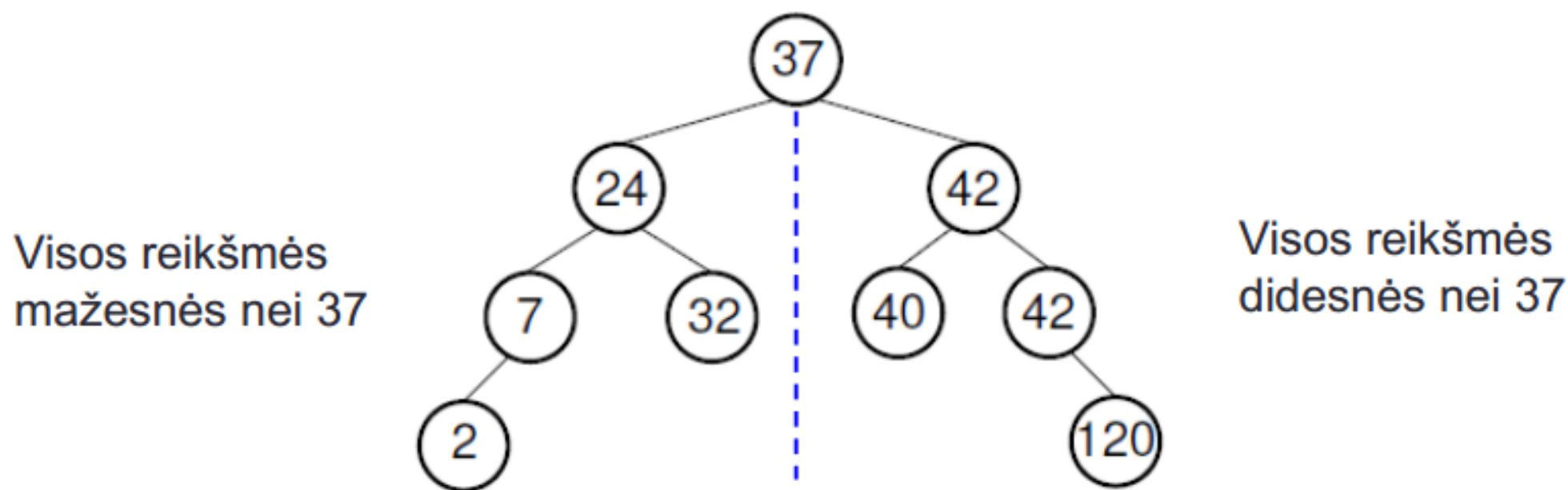
a + b / c * d - e * f

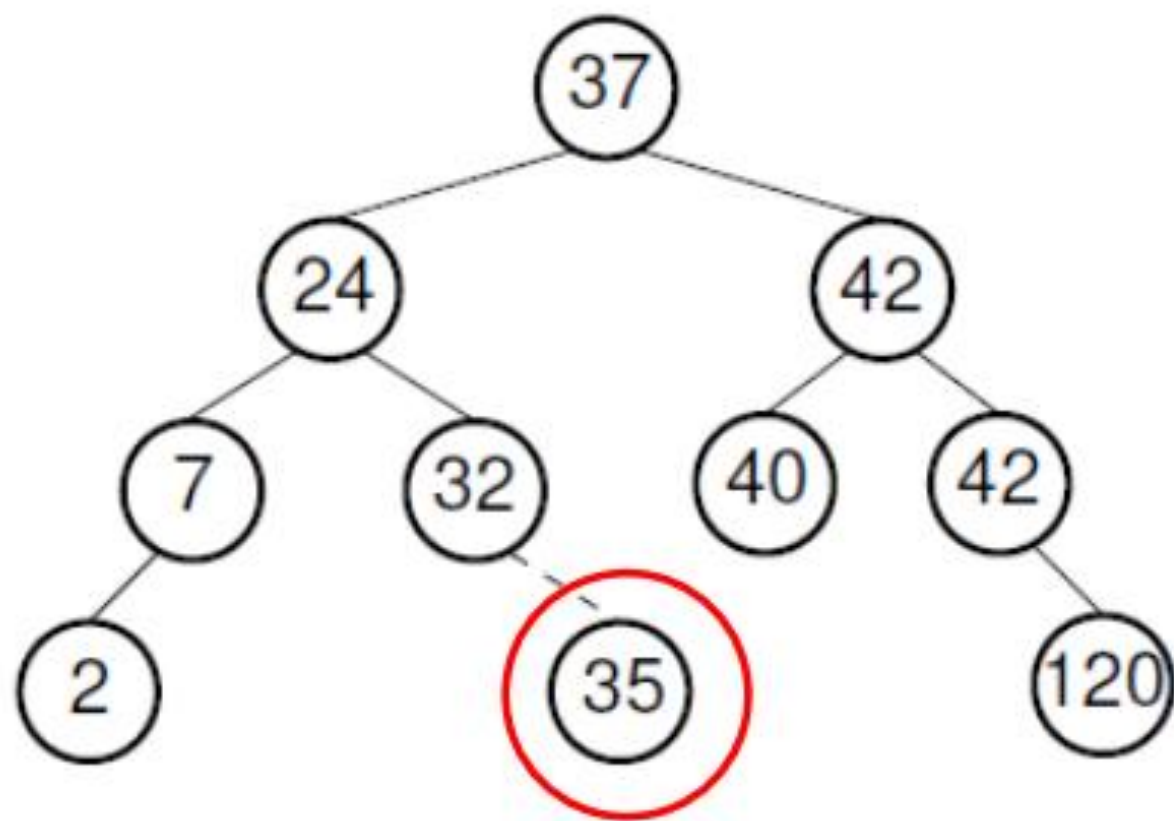
3. Atvirkštinis:

a b c / + d e f * - *

Dvejtainis paieškos medis tenkina tokias sąlygas:

- Visos viršūnės esančios kairėje šakoje nuo šaknies turi mažesnes reikšmes nei šaknis.
- Visos viršūnės esančios dešinėje šakoje nuo šaknies turi didesnes arba lygias reikšmes šakniai.

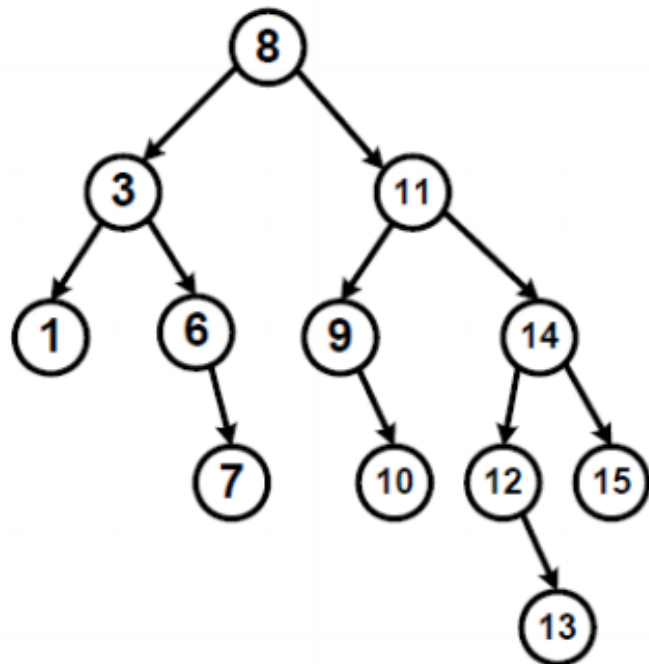




Iterpiama nauja viršūnė su reikšme 35.

DVEJETAINIO PAIEŠKOS MEDŽIO FORMAVIMAS

Pradiniai duomenys: 8, 11, 9, 3, 1, 14, 6, 12, 10, 7, 13, 15

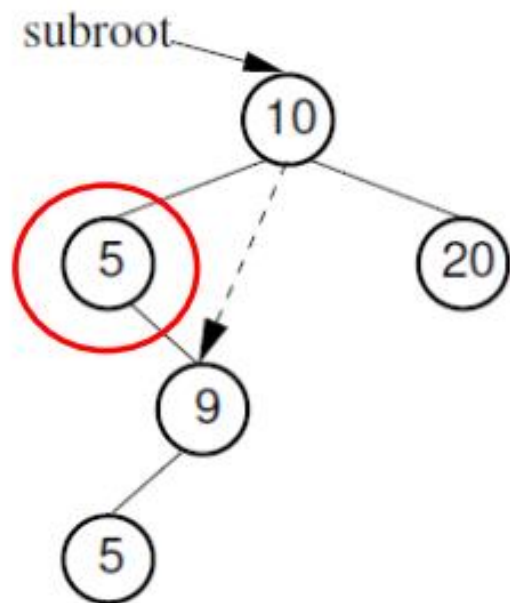


Formavimo principas:

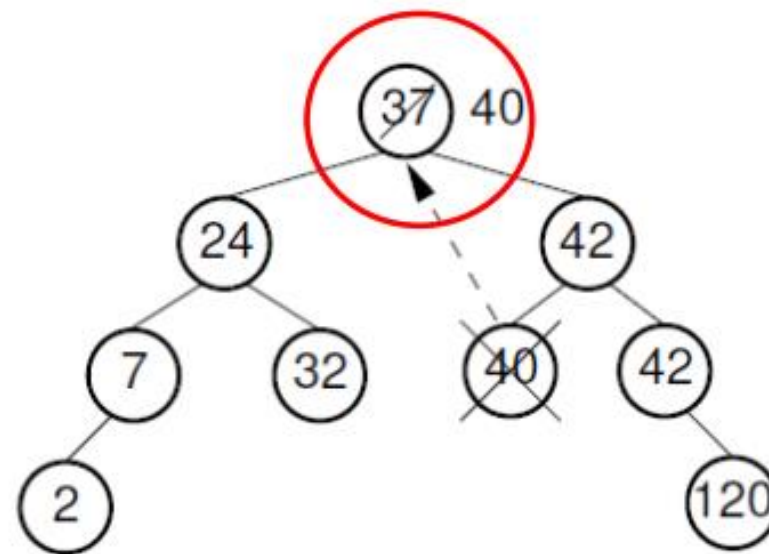
1. Pirmas elementas p_1 – šaknis.
2. Kitų elementų viršūnės šaka pasirenkama pagal nelygybės $p_i < p_k$ atsakymą:
 - teisybė – kairė šaka,
 - neteisybė – dešinė šaka.

Lyginimas pradedamas nuo šaknies.

VIRŠŪNĖS ŠALINIMAS



Jei trinama šakos viršūnė, ji keičiama į tos viršūnės dešinės šakos vaiką.



Jei trinama šaknis, jos reikšmė keičiama į dešinės šakos mažiausią reikšmę turinčią viršūnę.

SUTVARKYTAS MEDIS

Viena iš svarbių susijusių su algoritmais sąvokų, yra sutvarkytas medis. Medis sutvarkytas, kai jo viršūnėms priskirta tam tikra numeracija (eilės tvarka). Medžio viršūnės gali būti numeruojamos skirtingais būdais (paprastai nuo jų priklauso algoritmo efektyvumas).

MEDŽIO NUMERACIJA

Medžio numeracija gali būti žinoma, t.y. nurodyta tvarka, kuria vardijamos (ar spausdinamos, ar peržiūrimos) medžio viršūnės. Algoritmuose ar programavime aptinkamos tokios numeracijos (apibrėžimai rekursyvūs ir pateikti dvejetainiam medžiui):

MEDŽIO NUMERACIJA

infiksinis medis (angliškai infix arba inorder) - medžio viršūnės vardijamos taip: kairysis pomedis, viršūnė, dešinysis pomedis;

prefiksinis medis (angliškai prefix arba preorder) - medžio viršūnės vardijamos taip: viršūnė, kairysis pomedis, dešinysis pomedis;

postfiksinis medis (angliškai postfix arba postorder) - medžio viršūnės vardijamos taip: kairysis pomedis, dešinysis pomedis, viršūnė.

Šie apibrėžimai gali būti lengvai apibendrinti n-arinams medžiams. Jie yra analogiški algebrinių ar loginių reiškinių infiksinėms, prefiksinėms, postfiksinėms išraiškoms.

Infiksinė reiškinių išraiška - tai įprastas reiškinių užrašymas su skliausteliais: $((A + B) * (C - D) + E) * F$

Prefiksinė išraiška rašoma (rekursyviai) - operacijos ženklas, pirmas operandas, antras operandas: $* + * + A B - C D E F$

Postfiksinė išraiška rašoma (rekursyviai) - pirmas operandas, antras operandas, operacijos ženklas: $A B + C D - * E + F *$

MEDŽIO NUMERACIJA

Prefiksinėms ir postfiksinėms reiškinių išraiškoms skliausteliai gali būti nenaudojami. Postfiksine išraiška dar kitaip vadinama atvirkštine lenkiška notacija (pagerbiant lenkų logikus matematikus, įvedusius tokias išraiškas). Visų šių išraiškų atitikmenys gali būti dvejetainiai medžiai (infiksiniai, prefiksiniai, postfiksiniai), tačiau tik jų elementų numeravimo, o ne skaičiavimo prasme.

MEDŽIO NUMERACIJA

Dar vienas galimas medžių tvarkymo būdas - viršūnių numeracija pagal lygius: nulinio lygio viršūnė (šaknis); visos pirmo lygio viršūnės (numeruojamos iš kairės į dešinę); visos antro lygio viršūnės (taip pat iš kairės į dešinę); ir t.t.

Medžio struktūra yra iš esmės rekursyvi. Galima pateikti tokį rekursyvų medžio apibrėžimą:

Medis yra arba viršūnė, arba šakninė viršūnė, sujungta briaunomis su medžių aibe. Elementarios, tačiau naudingos medžio savybės yra tokios: 1. Medis su N viršūnių turi $N-1$ briaunų; 2. Dvejetainis medis su N vidinių viršūnių, turi $N+1$ išorinių viršūnių; 3. Dvejetainio medžio su N vidinių viršūnių išorinio kelio ilgis yra $2N$ didesnis už vidinio kelio ilgį; 4. Pilno dvejetainio medžio aukštis n su N vidinių viršūnių gali būti įvertinamas skaičiumi $\log_2 N$: $2^{n-1} < N + 1 \leq 2^n$.

Reiškinio $((A + B) * (C - D) + E) * F$ sintaksinis medis

