



Eötvös Loránd Tudományegyetem

Informatikai Kar

Komputeralgebra Tanszék

Prímszita algoritmusok összehasonlítása

Vatai Emil
Adjunktus

Nagy Péter
Programtervező Informatikus BSc

Budapest, 2018

Tartalomjegyzék

1. Bevezetés	3
2. Felhasználói dokumentáció	4
2.1. A megoldott feladat	4
2.2. Felhasznált módszerek	4
2.3. A program telepítése és futtatása	5
2.4. Prímek keresése	6
2.5. Mintaadatbázis karbantartása	7
2.6. Szitatóablák ellenőrzése	9
2.7. Sziták ellenőrzése	9
2.8. Minta megjelenítése	10
2.8.1. Minta hozzáadása	11
2.8.2. Minta közelítése	11
2.9. Sziták mérése scriptekkel	12
3. Fejlesztői dokumentáció	14
3.1. A feladat leírása	14
3.2. A program komponensei	15
3.3. A forráskód felosztása	15
3.4. Adatszerkezetek	15
3.5. Numerikus algoritmusok	15
3.6. Sziták	15
3.7. Prioritásos sorok	16
3.7.1. Bináris kupac	16
3.7.2. Bigyó	16
3.8. Memória	18
3.9. Teszt	18
3.9.1. Adatszerkezetek	18
3.9.2. Numerikus pontosság és sebesség	18
3.9.3. Sziták	18
3.9.4. Elmélet vs. mért	18

3.10. Források	18
--------------------------	----

1. fejezet

Bevezetés

Sziták összehasonlítása azonos környezetben. Hatékonyság összehasonlítása az elmélet szerint várható értékekkel. Ellenőrzésként a prímek néhány statisztikájának összevetése az elméleti értékekkel, és az ismert eredményekkel. Hatékonyság és implementálhatóság.

2. fejezet

Felhasználói dokumentáció

A megoldott feladat

A program lehetővé teszi különböző sziták futási idejének ábrázolását grafikonon, és összevetését elméleti becsült értékekkel. A programmal egyéb forrásból származó minták megjelenítése is lehetséges.

A prímszámok statisztikáinak előállításához a programhoz tartozik egy optimalizált szita-implementáció, amivel 2^{64} -ig lehet szitálni, és az eredményt rögzíteni. Az elmentett eredményeket a program egy külön része összesíti néhány alapstatisztikákra.

Különböző szita-algoritmusok futási idejének összehasonlításához a program tartalmazza Atkin szitájának egy implementációját, és Eratoszthenész szitájának néhány variációját.

Felhasznált módszerek

A program két rész fő részből áll. A prímszámok generátora C-99 nyelven készült el. A program csak parancssorból futtatható, és a sztenderd könyvtári függvényekből is csak néhányat használ a memóriájában előállított eredmények fájlba írásához. A C nyelv választását a hatékony végrehajtás és memórafelhasználás, és a hordozhatóság indokolja, a generátor elkülönítését az automatizálhatóság indokolja.

A prímgenerátor Eratoszthenész szitájának szegmentált változata, a szitatábla egy 2^{30} hosszú darabját szitálja ki minden iterációban. A generátor önmaga is két részből áll. Egy egyszerűbb implementáció a prímek listáját állítja elő 2^{32} -ig.

A gyorsabb, de összetettebb szita 2^{32} -tól $2^{64} - 2^{34}$ -ig szital, és a futásához szükséges a prímek listája 2^{32} -ig. A szitatáblát a memóriában a gyorsítótár hatékonyabb a kihasználásához kisebb részegemensekre osztja, és a prímeket a nagyságuk és a következő szitalási pozíciójuk szerint csoportosítja. Továbbá nagyobb prímeknek

csak harminccal relatív prím többszöröseit veszi figyelembe.

A program többi része a Java környezethez készült, a képzésben betöltött szerepe és a hordozhatósága miatt. A minták és függvények ábrázolásához használt Swing grafikus könyvtár a Java környezet része, egyszerű, hordozható, támogatja az élsimítást, és a sebessége elfogadható a feladatra. A Java sztenderd konténerosztályai habár kényelmesen használhatóak, és változatosak, de sebesség és memóriakihasználtsági problémák miatt több helyen is alacsonyabb szintű megoldást kellett választani.

A program a mintákat alapfüggvények lineáris kombinációjával képes közelíteni, a legkisebb négyzetek módszerével. A minták elemeinek nagy száma, és a mintaértékek nagy terjedelme miatt szükséges volt a numerikus algoritmusok nagyobb pontosságú implementációja, ami a számítási idő növekedésével jár.

A program hosszabb távon egyedül a prímek szegmensenkénti statisztikáit tárolja, amit csak sorban dolgoz fel, és a statisztikák összmérete sem indokolna egy egyszerű bináris fájlnál bonyolultabb megoldást. Az adatvesztés elkerüléséhez elég a statisztikafájl csere alapú felülírása.

A Java program tartalmaz több prímszita implementációt is, ezeknek az algoritmikus bonyolultsága különböző, de az összehasonlíthatóságához a közös részek közös implementációt kaptak, és az eltérő részek hasonló szinten optimalizáltak. A sziták helyességének ellenőrzéséhez a sziták eredménye egymással összehasonlítható. A programhoz elkészült Atkin szitája, és Eratoszthenész szitájának több variációja: az algoritmus legegyszerűbb implementációja a teljes szitatáblán jelöli meg a prímek többszöröseit, a többi a táblát rövidebb szegmensekre osztja fel, és ezeket sorban szitálja. A szegmentálás szükségessé teszi, hogy eldöntsük, hogy mely szegmensben mely prímek szitálnak. A legegyszerűbb megvalósítás minden szegmensben minden prímmel megpróbál szitálni, a próbaosztáshoz hasonlóan. Ennél hatékonyabb a két elsőbbségi sorral megvalósított szita, és a harmadik, edényekbe csoportosító megoldás.

A sziták között kapott helyet a próbaosztás, és az erős pszeudo-prím teszt is, de ezek a lassúságuk miatt csak igen rövid tartományok ellenőrzésére alkalmasak.

A program telepítése és futtatása

A program futtatásához Windows vagy Linux operációs rendszer szükséges, legalább 4GB szabad memóriával. A Java program futtatásához JRE8 szükséges, a fordításához JDK8. A C program fordításához C99 szabványú fordítóprogram kell. A program fejlesztése és tesztelése OpenJDK8-cal és GCC 7.3-mal történt Linux operációs rendszeren.

A tarball a Java programot lefordítva is tartalmazza, kicsomagolás után azonnal futtatható. A C programot a generator könyvtárban található Makefile segítségével lehet lefordítani:

```
generator$ make clean build
```

A Java program a NetBeans 8-as verziójával készült, parancssorból a project ant fájljának segítségével lehet újrafordítani:

```
gui$ ant clean jar
```

A program futtatásához több előkészített script is rendelkezésre áll, ezeket mind a scripts könyvtárból lehet indítani. Az előző két fordítást egyben is el lehet végezni a recompile scripttel:

```
scripts$ ./recompile
```

Prímek keresése

A prímek keresését két program végzi, az init program 2^{32} -ig keresi meg és tárolja el a prímeket, a generator 2^{32} -től $2^{64} - 2^{43}$ -ig. Az init készítette szitatáblák a generator futásához szükségesek.

Mindkét program a számokat 2^{30} hosszú táblánkként szitálja, minden tábla első száma $2^{30}k + 1$ alakú. A szitatáblákat a programok a fájlrendszerben bittérképként tárolják, egy tábla kb. 64Mb.

Az első négymilliárd szám szitálása a következő képpen indítható el:

```
generator$ ./init.bin ../db
szegmens 0 - kezdet          1 - felkészülés 159 906 063 ns - szitálás 1 122 878 403 ns
szegmens 1 - kezdet 1 073 741 825 - felkészülés 2 353 698 ns - szitálás 1 171 101 336 ns
szegmens 2 - kezdet 2 147 483 649 - felkészülés 2 344 515 ns - szitálás 1 188 330 478 ns
szegmens 3 - kezdet 3 221 225 473 - felkészülés 2 391 660 ns - szitálás 1 199 900 603 ns
```

Ennek eredmény 4 bittérképfájl, amit a db mappába ment.

A generator programot több féle képpen is lehet indítani. Mindegyik esetben két paramétert kell megadni az indításhoz, a számot, ahol szitálást kezdje, és a szegmensek számát, amit ebben a futásban végig kell szitálni. A kezdő számot meg lehet adni decimálisan is, vagy hexadecimálisan is, "0x" prefixszel. A szegmensek számát is kétféle képpen lehet szabályozni. Az egyik lehetőség fix számú szegmens megadása, decimálisan. A másik mód a háttértáron fenntartandó szabad hely átadása, decimálisan, vagy "0x" prefixszel hexadecimálisan.

```
generator$ ./generator.bin ../db start 0x100000001 segments 3
szegmens kezdet: 4 294 967 297
szegmensok száma: 3
kis szegmensok mérete: 22
elágazások bitjei: 8
felkészülés 1/4
felkészülés 2/4
felkészülés 3/4
felkészülés 4/4
felkészülés vége
szegmens 4 - kezdet 4 294 967 297 - felkészülés 8 162 031 487 ns - szitálás 729 914 370 ns
szegmens 5 - kezdet 5 368 709 121 - felkészülés          418 ns - szitálás 736 412 117 ns
szegmens 6 - kezdet 6 442 450 945 - felkészülés          385 ns - szitálás 750 022 409 ns
összes szitálás 2 216 348 896 ns
```

```
generator$ ./generator.bin ../db start 0x100000001 reserve-space 1000000000
szegmens kezdet: 4 294 967 297
fenntartandó hely: 1 000 000 000 byte
kis szegmensok mérete: 22
elágazások bitjei: 8
felkészülés 1/4
felkészülés 2/4
```

```

felkészülés 3/4
felkészülés 4/4
felkészülés vége
szegmens 4 – kezdet 4 294 967 297 – felkészülés 8 162 031 487 ns – szítálás 729 914 370 ns
szegmens 5 – kezdet 5 368 709 121 – felkészülés 418 ns – szítálás 736 412 117 ns
szegmens 6 – kezdet 6 442 450 945 – felkészülés 385 ns – szítálás 750 022 409 ns
összes szítálás 2 216 348 896 ns

```

Mindkét program kimenetében a start a szegmens kezdőszáma, a segment az indexe, $start = segment \cdot 2^{30} + 1$, az init a szita inicializálási ideje a szítálás előtt, és a sieve a szegmens szítálásának ideje. Ezek az információk az elmentett táblafájlokból is kiolvashatóak.

Mintaadatbázis karbantartása

A mintaadatbázis a lementett szitatáblák összesített statisztikáit tárolja. A program az adatbázis könyvtárában kétfajta fájlt vesz figyelembe. A "primes\[0-9a-f]{16}" reguláris kifejezésnek megfelelő nevű fájlok a szegmensek szitatáblái. A fájlnev második fele a szegmens kezdőszáma hexadecimálisan. A fájl a páratlan számok bit-térképe mellett tartalmazza a szegmens szítálásának megkezdésére, és a szítálására fordított időt, valamint ellenőrzésképpen a szegmens kezdőszámát is.

A szitatáblák mellett az adatbázis tartalmazhatja szegmensek összesített statisztikáit is, "aggregates" néven. Ez a fájl az eddig összesített szegmensről a következő információkat tartja nyilván:

- a szegmens kezdőszáma
- a szegmens szítálására való felkészülés idejét
- a szegmens szítálásának idejét
- az összesítésre fordított időt
- a szegmensfájl utolsó módosítási idejét
- a szegmensbe eső legkisebb, és legnagyobb prímet
- a szegmensbe eső prímek számát, $12\mathbb{Z}11$, $4\mathbb{Z}1$, $4\mathbb{Z}3$, $6\mathbb{Z}1$ alakok szerinti bontásban
- a szegmensben előforduló prímhézagok első előfordulásának helyét, és az előfordulások számát.

Az adatbázison három művelet végezhető:

- Le lehet kérni, hogy eddig hány szegmens van összesítve, és melyik a következő hiányzó szegmens. Azt is visszaadja, hogy hány szegmensfájl van épp az adatbázis könyvtárában, és hogy ezek között van-e olyan, ami nincs összesítve.

- Összesíteni lehet az új szegmensfájlok.
- Össze lehet fésülni két összesítőfájlt.

Mindhárom művelet elvégezhető grafikus felületen, és automatizáláshoz parancssorból is. A grafikus felületet a gui scripttel lehet indítani, a parancssorból a database script segítségével lehet elérni a műveleteket.

Ezek a scriptek határozzák meg, hogy a program melyik könyvtárat használja adatbázisként, ezt nem kell külön megadni. Szükség esetén ez a könyvtár a scriptek egyszerű módosításával megváltoztatható.

A szegmensinformációkat a grafikus felületen a "DB info" gomb megnyomásával lehet lekérdezni, parancssorból a "database info" kiadásával.

```
scripts$ ./database info
100.0%
szegmensfájlok: szegmensek száma: 7
szegmensfájlok: első szegmens kezdete: 1
szegmensfájlok: utolsó szegmens kezdete: 6,442,450,945
szegmensfájlok: első hiányzó szegmens kezdete: 7,516,192,769
szegmensfájlok: hiányzó szegmensek száma: 18,446,744,073,709,551,593

összesítések: szegmensek száma: 62,881
összesítések: első szegmens kezdete: 1
összesítések: utolsó szegmens kezdete: 67,516,885,893,121
összesítések: első hiányzó szegmens kezdete: 67,517,959,634,945
összesítések: hiányzó szegmensek száma: 18,446,744,073,709,488,719

új szegmensfájlok: 7

scripts$ ./database info crunch 100
100.0%
szegmensfájlok: szegmensek száma: 7
szegmensfájlok: első szegmens kezdete: 1
szegmensfájlok: utolsó szegmens kezdete: 6,442,450,945
szegmensfájlok: első hiányzó szegmens kezdete: 7,516,192,769
szegmensfájlok: hiányzó szegmensek száma: 18,446,744,073,709,551,593

összesítések: szegmensek száma: 62,881
összesítések: első szegmens kezdete: 1
összesítések: utolsó szegmens kezdete: 67,516,885,893,121
összesítések: első hiányzó szegmens kezdete: 67,517,959,634,945
összesítések: hiányzó szegmensek száma: 18,446,744,073,709,488,719

crunch: ../generator/generator.bin ../db start 0x3d6840000001 segments 0x64
```

A "crunch" utótag és szitálni kívánt szegmensek számának megadásával az utolsó sorban a következő futtatásra ajánlott parancsot is megjeleníti a program. Ezeknek a parancsoknak a követésével a teljes szitálható tartományt fel lehetne dolgozni. A "crunch-numbers" script ezt a feladatot kísérli meg elvégezni.

Az info megjeleníti az új szegmensfájlok számát is, azokat, amikhez vagy nem tartozik összesített adat, vagy tartozik, de az ott tárolt utolsó módosításnál újabb a szegmensfájl. Ezeket a "DB összesítés" gomb megnyomásával, vagy a "database reaggregate" paranccsal lehet összesíteni.

```
scripts$ ./database reaggregate
100.0%
```

Lehetőség van egy másik adatbázisból statisztikák átvételére is. Az összefésülést a grafikus felületen a "DB import" megnyomása után egy file kiválasztásával lehet indítani, parancssorból a "database import aggregates fájlnev" kiadásával. A megadott

fájlból azokat a szegmeneseket veszi át a program, amik vagy nincsenek még meg az adatbázisban, vagy megvannak, de az importált statisztikák utolsó módosítási ideje nagyobb az adatbázisban lévőnél.

```
scripts$ ./database import aggregates ../db/aggregates.large
100.0%
```

screenshot

Szitatáblák ellenőrzése

A generator program teszteléséhez a grafikus program képes a szegmensfájlok ellenőrzésére. Ehhez a program újra előállítja az ellenőrzött szegmensfájlokat, és azokat összeveti. A újraelőállításához egy egyszerű szegmentált Eratoszthenészi szitát, vagy ez erős pszeudo-prím tesztet lehet választani. A pszeudo-prím teszt rendkívül lassú, a célja, hogy a szitákhoz képest egy alapvetően más módszer is rendelkezésre álljon az ellenőrzéshez.

A grafikus felületen a "Szegmensfájlok ellenőrzése" gomb megnyomása után lehet kiválasztani a szitálni kívánt szegmenseket, és a referencia előállításának módszerét, majd az "Ellenőrzés" gombbal lehet megkezdeni a tényleges folyamatot.

Az ellenőrzést parancssorból is lehet futtatni a check-segments scripttel. A script első paraméterében a referencia módszerét várja, ami lehet "sieve", vagy "test", a paraméterek maradéka az ellenőrizendő szegmensek kezdőszáma kell legyen. Ha egy kezdet nincs megadva, akkor a program az adatbázis könyvtárban található összes szegmensfájlt ellenőrzi.

```
scripts$ ./check-segments sieve
100.0%: A szegmensek helyesek.

scripts$ ./check-segments sieve 0x1 0xc0000001
100.0%: A szegmensek helyesek.

scripts$ ./check-segments test 0x1
100.0%: A szegmensek helyesek.
```

Ha a kezdőszám nagyobb, mint 1, a szitával ellenőrzéshez szükséges, hogy az init program által előállított négy darab szegmensfájl tartalmazza az adatbáziskönyvtár, mert a szita kezdőszámra pozícionálásához a prímszámokat innen olvassa fel a program.

screenshot

Sziták ellenőrzése

Az sebességek összehasonlításához írt sziták eredményét is lehet ellenőrizni, az összehasonlításához itt is a szegmensek ellenőrzésénél használt egyszerű szegmentált szita a referencia.

A program grafikus felületén a "Sziták ellenőrzése" lehetőséget választva ki kell választani az ellenőrizendő szitát, és az ellenőrizni kívánt intervallum kezdő és végző számát. A folyamat az "Ellenőrzés" gomb megnyomásával indítható.

Parancssorból is ellenőrizhetők a sziták a "check-sieve" scripttel. A script 3 paramétere a szita neve, és az intervallum kezdő-, és végszáma. Az ellenőrizhető sziták nevei:

- atkin: Atkin szitája
- cols: Cache optimized linear sieve
- eratosthenes-segmented: Eratoszthenész szitájának szegmentált változata
- bin-heap és bin-heap-inplace: Eratoszthenész szitája bináris kupaccal, az bin-heap-inplace a kupac javítását helyben végzi
- buckets, buckets-n, buckets-simple: Eratoszthenész szitája edényekkel, a buckets-simple nem szegmentált, a buckets, és a buckets-n szegmentált. a buckets-n-ben az n helyére 1-től 8-ig a számrendszer bitjeit lehet megadni, pl. buckets-3 8as számrendszerben számol
- trial-division: próbaosztás

A sziták nevei listázhatóak a list-sieves scripttel, a describe-sieves script rövid leírást is ad.

```
scripts$ ./check-sieve buckets-3 30000000000001 30100000000001  
100.0%: A szita helyes.
```

Ha a kezdőszám nagyobb, mint egy, az ellenőrzéshez ilyenkor is szükséges, hogy a négy darab kezdő szegmensfájlt tartalmazza az adatbáziskönyvtár.

[screenshot](#)

Minta megjelenítése

A program a grafikus felületén képes mintákat ábrázolni, a mintákat függvényekkel közelíteni, és ezeket a függvényeket is ábrázolni, valamint kiszámolja a közelítések négyzetes hibáját a mintához képest.

A grafikonábrázoló részt a gui-graph scripttel lehet indítani, vagy a gui scripttel, majd a "Grafikon" gomb megnyomásával.

Az ablak bal oldalán jelennek meg a betöltött minták és közelítő függvények grafikonjai. Minta és függvény hozzáadása után a grafikon megjelenített részét a program úgy választja meg, hogy minden minta összes pontja ebbe a nézetbe essen. A nézet a grafikon feletti gombokkal, vagy egérrel mozgatható, nagyítható,

vagy kicsinyíthető. Az "Auto.nézet" gombbal a nézet a minták teljes terjedelmére visszaállítható.

Az ablak jobb oldala a betöltött minták kezelésére szolgál. A felső harmada az összes betöltött mintát mutatja. Az ablak jobb oldalának alsó két harmadában az itt kiválasztott minta részletei látszódnak.

Minta hozzáadása

A "Minta hozzáadásánál" három adatforrásból lehet választani. A "Fájl betöltésével" CSV fájl olvasható be. A fájl formátuma megengedő, az első oszlopot x, a második oszlopot y értékeknek próbálja értelmezni, a fel nem ismert sorokat eldobja. Ha az első sor első oszlopa "BARS", akkor vonaldiagram helyett a mintát oszlopdiaagramként fogja ábrázolni.

A "Prímstatisztikák" a generator kimeneteléből összesített statisztikák, amiket az adatbáziskönyvtárban keres. A program a szegmensstatisztikák hiánytalan kezdőszeletét veszi csak figyelembe. A statisztikáknak négy csoportja van. A prímek száma, és néhány kiemelt alakú prím száma a szegmensek végén. Ide tartozik ugyanezen a helyeken a prímszámtétel becslése, és a becslés abszolút és relatív hibája is. Az ikerprímek száma a szegmensek végén. A prímhézagok statisztikái a szegmensek teljes kezdőszeletére vonatkozik. Ezek az előfordult hézagok gyakorisága, első előfordulása, jósága, és az legnagyobb prímhézag adott számig. Ezeken kívül a szegmensek szitálásának ideje is betölthető. Az összesített idő a szegmensek végig az összes addigi szegmens összes ideje, a szegmensenkénti a szegmensek végéhez az adott szegmens szitálásának idejét rendeli.

A minták harmadik forrása a programba épített sziták futási idejének mérése. Ehhez meg kell adni a szitát, a szitált intervallum elejét és végét, a szita belső szegmensméretét, a generálandó minták számát, és a mérések számát, a mértéket, és hogy az eredményt összesítse-e. A sziták listája ugyanaz, mint a sziták ellenőrzésénél. A mintákat a program a szitált intervallumon egyenletesen osztja el a megadott méretű szegmensek határainál. A teljes szitálást a megadott mérések számszor végzi el, és az eredményt átlagolja. Összesítést kiválasztva a minta értéke a szitált intervallum elejétől az adott számig szitálás teljes ideje, különben az előző minta helyétől. A mérték szabályozza, hogy a tényleges eltelt időt mérje a program nanoszekundumban, vagy az elvégzett műveletek számát.

screenshot

Minta közelítése

Az ablak jobb oldalának középső harmadában a kiválasztott minta részletei látszódnak. Itt változtatható meg a minta ábrázolásához használt szín, távolítható el a

minta, vagy menthető CSV fájlba.

A táblázat a minta közelítéseit mutatja, itt olvashatóak le a közelítés komponensfüggvényei, és a közelítés hibája. A "Közelítés" gombbal lehet új közelítést hozzáadni az ábrázoláshoz. A közelítéshez a lineáris legkisebb négyzetek módszerét használja a program, ehhez a lineáris kombináció elemi függvényeit meg kell adni. A függvények kiválaszthatóak az előre megadott listából, vagy javascript nyelven is megadhatóak. A közelítéshez csak olyan függvények használhatóak, amik a minta minden pontjában értelmezve vannak.

Javascript függvényhez meg kell adni a függvény nevét és forráskódját. A forráskódban a függvény argumentumára az x változóval lehet hivatkozni, és a script utolsó kifejezésének értéke lesz a függvény értéke x -ben. Pl. az $x \mapsto \frac{\ln x}{x}$ hozzárendelés forráskódja:

```
Math.log(x)/x
```

A scriptben használható változó, és a szokásos vezérlőutasítások. Pl. az iterált logaritmus egy megadása:

```
var y=0;
while (1<x) {
  x=Math.log(x);
  ++y;
}
y
```

Az így megadott függvények visszatérési értéke mindig double kell legyen, és a függvény azokban a pontokban értelmezett, ahol ez az érték véges.

Egy kiválasztott közelítés részletei az ablak jobb alsó sarkában látszódnak, itt megváltoztatható a közelítéshez rendelt szín, és eltávolítható a közelítés. A táblázat a közelítéshez kiválasztott függvények együtthatóit mutatja a lineáris kombinációban.

[screenshot](#)

Sziták mérése scriptekkel

A sziták futási idejének mintáit parancssorból is elő lehet állítani a measure-sieve scripttel, paraméterként meg kell adni ebben a sorrendben:

- a szita nevét, ezek ugyan azok lehetnek, mint a check-sieve scriptnél
- a számot, ahol a szitálás kezdődik, ez páratlan szám kell legyen
- a szitálás végét, ez páratlan szám kell legyen

- a szegmensek méretét
- a mérések számát
- a minták számát
- hogy időt (nanosecs), vagy műveleteket (operations) mérjen
- összesítse az időket ("sum"), vagy szegmensenként külön számolja ("segment")
- a fájlt, ahova az eredményt mentse CSV-ben

A számokat meg lehet adni decimálisan, és hexadecimálisan is "0x" előtaggal. Ha a kezdőszám nagyobb, mint 1, akkor az adatbáziskönyvtárnak tartalmaznia kell az init program generálta első négy szegmenst.

Lehetőség van a mérték és az összesítés mind a négy kombinációját egy méréssel előállítani, ehhez mind a négy kimeneti fájl meg kell adni külön.

```
scripts$ ./measure-sieve buckets-3 1 0x10000001 0x10000 3 1000 nanosecs sum out1.csv
100.0%
scripts$ ./measure-sieve buckets-3 1 0x10000001 0x10000 3 1000 operations segment out2.csv
100.0%
scripts$ ./measure-sieve buckets-3 1 0x10000001 0x10000 3 1000
seg-ns.csv seg-ops.csv sum-ns.csv sum-ops.csv
100.0%
```

A mérések megkönnyítéséhez a measure-sieves script az összes szítét megméri, és az eredményt a samples könyvtárba menti.

teljes. generator. atkin szegmens méret.

3. fejezet

Fejlesztői dokumentáció

A feladat leírása

A program fő feladata, hogy lehetővé tegye prímszíták futási idejének összehasonlítását. Ehhez keretet kell adjon, hogy azonos paraméterekkel lehessen a szíták futási idejét mérni, és egyszerű, egymáshoz hasonlóan optimalizált szítákat kell implementálni, hogy a futási időket a fő algoritmusok határozzák meg, mivel a szítálást felgyorsító egyes ötletek nem minden szítán alkalmazhatóak. Elvárás, hogy a futási időt eltelt időben, és összeszámolt műveletekben is lehessen mérni.

A szítákat szegmentáltan kell megvalósítani, és szabályozhatóvá kell tenni a szegmensek méretét, és a szítálás kezdő és végszegmensét. A megvalósítandó szíták:

- Atkin szítája
- Eratoszthenész szítája
 - egyszerű tömbbel
 - bináris kupaccal
 - edényekkel
 - edények monoton sorával

A programnak grafikusán ábrázolnia kell a futási időket, és lehetőséget kell adjon ezek közelítésére függvényekkel. A közelítéshez a legkisebb négyzetek módszerét kell használni, és az elemi függvények legyenek megválaszthatóak. Elemi függvényt lehessen megadni JavaScript nyelven is.

A grafikus ábrázolás legyen képes külső forrásokból származó minták megjelenítésére és közelítésére is. A prímszámok néhány statisztikájának megjelenítéséhez elő kell tudni állítani ezeket a statisztikákat nagyobb intervallumokon is. Ehhez adni kell egy szítát, ami hatékonyan képes 2^{64} -ig szítálni, és ennek az eredményét összesíteni kell. Az előállítható statisztikák adott n -ig szítálva legyenek:

- a prímek száma
- a $4\mathbb{Z} + 1, 4\mathbb{Z} + 3, 6\mathbb{Z} + 1, 12\mathbb{Z} + 11$ alakú prímek száma
- az ikerprímek száma
- az előforduló prímhézagok első előfordulásának helye
- a prímhézagok előfordulásának száma
- a szitálásra fordított idő

A prímek számát, az ikerprímek számát, és a szitálás idejét nem csak a szitált intervallum végén kell előállítani, hanem az intervallum belsejében is több ponton.

A sziták sebességének mérését, és a statisztikák előállítását parancssorból is el kell tudni végezni, hogy szkriptekkel ez automatizálható legyen. A statisztikák előállításánál ezen túl még elvárás, hogy a folyamat leállítható, és később folytatható legyen.

A program komponensei

A forráskód felosztása

Adatszerkezetek

Numerikus algoritmusok

Egyenletrendszerek. Összeadás. Körül kéne írni, hogy igazán tudjuk, hogy hipotézisvizsgálatra nem vállalkozunk. A grafikonrajzolónak sokkal okosabbnak kéne lennie, hogy gyorsan változó függvényeket helyesen jelenítsen meg, reménytelenül okosnak.

Sziták

Eratosztenész szitája, szegmentáltan is. COLS. Prioritásos sorral. Atkin szitája.

Szegmentált szita inicializálása.

Trial division. Pszeudoprím teszt.

Feltételek. Elméleti sebesség.

Prioritásos sorok

```
1:  $q \leftarrow \text{ÚJ-SOR}$ 
2: for  $i \leftarrow 2, n$  do
3:   while  $\exists(p, k) \in q : k \leq i$  do
4:      $(p, k) \leftarrow \text{SOR-ELTÁVOLÍT-MIN}(q)$ 
5:      $\text{MEGJELÖL}(i)$ 
6:      $\text{SOR-BESZÚR}(q, (p, k+p))$ 
7:   end while
8:   if  $\neg \text{MEGJELÖLT?}(i)$  then
9:      $\text{SOR-BESZÚR}(q, (i, 2i))$ 
10:  end if
11: end for
```

Bináris kupac

A mérések grafikonják pixelei alapján lassú. A beszúrásonkénti elméleti $\mathcal{O}(\log|q|)$ ideje se biztató.

Bigyó

A bigyó egy monoton prioritásos sor. A sor monoton, minden állapothoz tartozik egy érték, a sor aktuális pozíciója, aminél kisebb vagy egyenlő pozíciójú értéket a sor nem tartalmazhat. A bigyó edények egy végtelen sorozatát is tárolja, a sor elemei ezekbe az edényekbe kerülnek. Egy eltárolt elem helyét a sorozatban az elem pozíciójának és a sor aktuális pozíciójának távolsága határozza meg.

A távolságfüggvény legyen

$$d(x, y) := \lfloor \log_2(x \oplus y) \rfloor \quad (x, y \in \mathbb{N}, y > x \geq 0)$$

ahol \oplus a bitenkénti XOR.

$d(x, y)$ a legnagyobb bit-index, ahol x és y eltér.

Ha q egy bigyó, legyen $q.a$ q aktuális pozíciója, és $q.e[i]$ q i . edénye. AZ edények, és a számpárok struktúrája...

Egy q bigyó invariánsa

$$\forall (p, k) \in q :$$

$$q.a < k$$

$$\forall i \in \mathbb{N}_0 : (p, k) \in q.e[i] \iff i = d(q.a, k)$$

$$\forall (p, k) \notin q : \forall i \in \mathbb{N}_0 : (p, k) \notin q.e[i]$$

Új, üres sor létrehozása tetszőleges kezdőpozíciótól, és meglévő sorba elem beszúrása...

A sor elemeinek feldolgozása i -ig

```
1: while  $q.a < i$  do
2:    $j \leftarrow d(q.a, q.a + 1)$ 
3:    $q.a \leftarrow q.a + 1$ 
4:   for all  $(p, k) \in q.e[j]$  do
5:     EDÉNY-KIVESZ( $q.e[j], (p, k)$ )
6:     if  $k = i$  then
7:       VISSZAAD( $(p, k)$ )
8:     else
9:       EDÉNY-BESZÚR( $d(q.a, k), (p, k)$ )
10:    end if
11:  end for
12: end while
```

Helyesség

Idő

Hely

Számrendszer

És számrendszer vs. tisztán funkcionálisban nincs bármekkora tömb, csak valami fával közelítve.

Amúgy sincs bármekkora tömb...

És exponenciálisan kell növelni...

Cache

És exponenciálisan kell növelni...

Memória

Összes prím, és pozíciója 2^{32} -ig. Primitív típusok és boxing. Garbage collector.

Teszt

Adatszerkezetek

Numerikus pontosság és sebesség

Sziták

Elmélet vs. mért

Források