

Circles graph

Uživatelská dokumentace

*Michal Drbohlav | [drbohmi1@fit.cvut.cz](mailto:drbohmi1@fit.cvut.cz)*

# Obsah

[Úvod](#)

[Popis programu](#)

[Popis činnosti](#)

[Formát vstupních dat](#)

[Výstup](#)

[Efekt](#)

[Ovládání programu](#)

[Konfigurační soubor](#)

[Nastavení](#)

[Timestamp formát](#)

[Seznam časových specifikátorů](#)

[Y-min](#)

[Y-max](#)

[X-min](#)

[X-max](#)

[Speed](#)

[FPS](#)

[Time](#)

[Legend](#)

[Gnuplot params](#)

[Effect params](#)

[Delay](#)

[Columns](#)

[Color](#)

[Method](#)

[Steps](#)

[Config file](#)

[Name](#)

[Ignore errors](#)

## Úvod

Jedná se o script primárně určený pro zpracování logů obsahujících nějakou časovou značku s hodnotou. Z těchto dat script vytvoří animaci grafu ve formátu **'mp4'**. Script se spouští z příkazové řádky a umožňuje rozsáhlou parametrizaci, která ovlivňuje výsledný efekt. Zdrojová data mohou být umístěna ve více souborech nebo mohou být zadána jako odkaz na web.

## Popis programu

Script se skládá z hlavního, spouštěcího souboru a poté ze souboru s funkcemi. Soubor **'circles\_graph.py'** je spustitelný soubor, který je hlavní částí scriptu. Tento soubor využívá ke svému fungování všechny fce ze souboru **'functions.py'**. Pro spouštění a ovládání se používá pouze tento soubor.

Soubor **'functions.py'** musí být pro správné fungování ve stejném adresáři jako soubor **'circles\_graph.py'**.

Dále je potřeba, aby byly nainstalovány programy **gnuplot** a **ffmpeg**. Pokud nebudou tyto programy nainstalovány, tak script nebude pracovat a při spuštění ohlásí chybu.

Script pracuje se zdrojovými soubory jejichž formát je popsán níže. Z těchto souborů vytváří výslednou animaci, ve které se snaží spojit zdrojové soubory nejvhodnějším možným způsobem, pokud je to možné.

## Popis činnosti

Vytvoření jedné animace ze sady vstupních dat je proces skládající se ze 4 částí. Při výchozím nastavení se aktuálně prováděná část a aktuální stav zpracování vypisují na terminál. Tento výpis lze upravit použitím parametru **'-v'**.

První fází je zpracování vstupních souborů. V této části se prochází postupně všechny vstupní soubory a kontrolují se, zda neobsahují chyby. Správný formát vstupních dat je popsán níže. Při výchozím nastavení se program ukončí ihned po detekci chyby. Použitím přepínače **'-E'** nebo direktivy **'IgnoreErrors'** v konfiguračním souboru lze nastavit ignorování chyb. V takovém případě se script po detekci chyby neukončí, ale pouze se přeskočí chybný řádek. V případě, že žádný zdrojový soubor neobsahuje řádek ve správném formátu, tak se na konci této části celý script ukončí. Doba vykonání této fáze je závislá na počtu a velikosti vstupních souborů.

Duhou fází je příprava dat. V této části script připravuje data pro výsledný efekt. Jelikož pracuje s daty zkontrolovanými v první fázi, tak již v této fázi nemůže dojít k chybám zapříčiněným špatným formátem obsahu vstupních souborů nebo nesprávným nastavením. Doba vykonání této fáze je závislá na počtu a velikosti vstupních souborů.

Třetí fází je generování grafu. V této fázi vytváří dočasně obrázky ve formátu **'png'**, které se v poslední fázi spojí do výsledné animace. Stejně jako v druhé fázi zde nemůže dojít k chybám zapříčiněným špatným formátem vstupních dat. Doba vykonání této fáze je závislá na nastavení **'columns'**, **'delay'**, **'time'**, **'speed'** nebo **'fps'**.

Poslední fází je generování animace. V této části se vytváří výsledná animace z obrázků, vytvořených v předchozí fázi. Název výsledného souboru je ve výchozím nastavení

**'circles\_graph.mp4'**. Název lze změnit použitím parametru **'-n'**, nebo direktivy **'name'** v konfiguračním souboru. Animace je uložena do adresáře se stejným názvem. Pokud takový adresář již existuje, tak se pro nový adresář použije název **'name\_i'** kde  $i = \max(0, i) + 1$ . Uživatel, který tento program spouští, musí mít v aktuální adresáři povoleno právo zápisu.

## Formát vstupních dat

Script vyžaduje, aby každý řádek zdrojových souborů měl formát: 'časová značka' 'hodnota'. Mezi časovou značkou a hodnotou může být libovolné množství mezer nebo tabelátorů.

Zdrojový soubor může také obsahovat libovolné množství prázdných řádků.

Časová značka musí odpovídat nastavení časového formátu (přepínač **'-t'**, nebo direktiva **'timeformat'**). Řádky v každém souboru musí být seřazeny vzestupně podle této časové značky. Hodnota je číslo typu float nebo int. Toto číslo může být kladné, nebo záporné.

Každý řádek, který neodpovídá tomuto formátu je považován za chybný.

## Výstup

Výstupem je animace ve formátu **'mp4'**. Ze všech zadaných vstupních souborů je vytvořena jedna animace podle efektu popsaného níže. V případě, že se datумы ve vstupních souborech překrývají, tak se pro každý soubor vygenerují vlastní značky, jinak se spojí do jedné sady. V případě, že se data nepřekrývají, tak se soubory seřadí podle datumů vzestupně. Barvy a jiné vlastnosti parametru lze změnit pomocí parametru **'-e'**, nebo direktivy **'effectparams'**. Více nastavení efektu je popsáno níže.

### Příklad spojení sady vstupních souborů do jednoho:

	soubor_1	soubor_2	soubor_3
první řádek	[10:00:00] 1	[10:06:00] 2	[10:05:30] 3
...	...	...	...
poslední řádek	[10:05:00] 1	[10:11:00] 2	[10:05:59] 3

**Příkaz:** `./circles_graph.py -t „[ %H:%M:%S ]“ soubor1 soubor2 soubor3`

Po zadání předchozího příkazu dojde ke spojení souborů v pořadí 'soubor\_1', 'soubor\_3' a 'soubor\_2'. Na hodnotách nezáleží, soubory se řadí pouze podle datumů.

Předpokládá se, že všechny vstupní soubory spolu nějakým způsobem souvisí, navazují na sebe, nebo se překrývají (i částečně).

Pro výslednou animaci je vytvořen nový adresář, do kterého je uložena, toto je blíže popsáno dále.

## Efekt

Efekt animace spočívá ve vytvoření sady bodů pro každý vstupní soubor, nebo pro více souborů spojených do jedné sady. Celý graf je ve výchozím nastavení rozdělen na 30 rozsahů (sloupců). Toto nastavení lze změnit pomocí parametru **'-e'**, nebo direktivy **'effectparams'** v konfiguračním souboru nastavením `'columns="hodnota"'`. Tato hodnota udává počet sloupců pro celý graf, ne pro jednotlivé vstupní soubory či jejich sady. Proto v případě, že některá sada sloupců nesahá přes celý graf, bude tato sada sloupců obsahovat menší počet sloupců. Graf je rozdělen na tolik časových rozsahů stejné velikosti, kolik tato

hodnota udává, a šířka každého sloupce odpovídá právě tomuto rozsahu. Výška sloupce (osa Y) je vypočtena ze všech hodnot jejichž časová značka v souboru odpovídá časovému rozsahu pro daný sloupec. Výpočet této hodnoty závisí na použité metodě výpočtu. Výsledný efekt zahrnuje postupné přibližování se bodů od krajů k žádané hodnotě. Body se zobrazují postupně zleva. Pomocí přepínače **'-e'**, nebo direktivy **'effectparams'** lze nastavit, o kolik kroků později má začít růst každý následující sloupec, a to pomocí `'delay="hodnota"'`. Veškerá parametrizace efektu je popsána v kapitole Nastavení.

## Ovládání programu

Program se ovládá pomocí přepínačů na příkazové řádce nebo pomocí konfiguračního souboru. Většina nastavení je nepovinná a při jejich nedefinování se používají výchozí hodnoty. Minimální nutné nastavení programu zahrnuje pouze definici zdrojových dat. Při nastavení programu pomocí přepínačů i pomocí konfiguračního souboru mají přednost přepínače. Většina přepínačů, až na výjimky popsané v kapitole 'Nastavení', lze použít pouze jednou. V konfiguračním souboru je možné použít stejné nastavení vícekrát. Pokud se nejedná o nastavení **'effectparams'** nebo **'gnuplotparams'**, tak program zohlední pouze poslední toto nastavení.

### Konfigurační soubor

Struktura konfiguračního souboru musí mít formát: **'direktiva' 'hodnota'**. Každá další direktiva a její hodnota musí být na novém řádku. Prázdné řádky nebo řádky obsahující pouze bílé znaky program ignoruje. Direktiva a její hodnota musí být oddělena mezerou nebo tabelátorem (případně jejich kombinací). Každý řádek začínající znakem **'#'** je celý považován za komentář. Veškerá direktiva a její správná hodnota je popsána níže v kapitole Nastavení. Program nehledí na velikost písmen použitých při definování direktivy. Při vícenásobném použití jedné direktivy (pokud se nejedná o direktivu **'effectparams'** nebo **'gnuplotparams'**) se zohlední pouze poslední definovaná hodnota. Každý řádek, který neodpovídá tomuto formátu, bude považován za chybný.

## Nastavení

Všechna nastavení, kromě nastavení config file, je možné definovat jak na příkazové řádce, tak v konfiguračním souboru. Nastavení config file lze definovat pouze na příkazové řádce. Následující kapitoly popisují veškeré nastavení scriptu a jejich definování na příkazové řádce nebo v konfiguračním souboru.

## Timestamp formát

Timestamp formát udává očekávaný časový formát všech vstupních souborů. Časový formát může být libovolný řetězec, který ale musí obsahovat minimálně jeden časový specifikátor.

### Seznam časových specifikátorů

%Y	rok na 4 cifry
%y	rok na poslední 2 cifry
%m	měsíc na 2 cifry s vedoucí nulou
%d	den na 2 cifry s vedoucí nulou
%H	hodiny
%M	minuty
%S	vteřiny

Každý časový specifikátor může být použit pouze jednou. Specifikátor '%Y' nemůže být použit současně se specifikátorem '%y'. Chybné zadání timestamp formátu je považováno za kritickou chybu a vede k ukončení programu bez ohledu na nastavení 'ignore errors'.

Timestamp formát může být nastaven pouze jednou.

K nastavení timestamp formátu na příkazové řádce slouží přepínač '-t', v konfiguračním souboru potom direktiva 'TimeFormat'.

#### Příklad nastavení timestamp formátu souboru 'data':

```
./circles_graph.py -t „[%Y/%m/%d %H:%M:%S]” data
```

**Konfigurační soubor:** TimeFormat [%Y/%m/%d %H:%M:%S]

**Výchozí nastavení:** [%Y-%m-%d %H:%M:%S]

## Y-min

Nastavení 'y-min' slouží k určení minimální zobrazované hodnoty grafu na ose Y. 'Y-min' může být nastaveno pomocí přepínače '-y', nebo pomocí direktivy 'Ymin' v konfiguračním souboru, a to pouze jednou. Přípustné hodnoty jsou 'min', nebo konkrétní hodnota. Toto nastavení je popsáno v následujících příkladech. Výchozí nastavení je 'min'.

**Příklad nastavení - 'min':** ./circles\_graph.py -y min data

**Konfigurační soubor:** Ymin min

**Příklad nastavení - 'hodnota':** ./circles\_graph.py -y 10 data

**Konfigurační soubor:** Ymin 10

## Y-max

Nastavení 'y-max' slouží k určení maximální zobrazované hodnoty grafu na ose Y. 'Y-max' může být nastaveno pomocí přepínače '-Y', nebo pomocí direktivy 'Ymax' v konfiguračním souboru, a to pouze jednou. Přípustné hodnoty jsou 'max', nebo konkrétní hodnota. Toto nastavení je popsáno v následujících příkladech. Výchozí nastavení je 'max'.

**Příklad nastavení - 'max':** `./circles_graph.py -Y max data`

**Konfigurační soubor:** `Ymax max`

**Příklad nastavení - 'hodnota':** `./circles_graph.py -Y 10 data`

**Konfigurační soubor:** `Ymax 10`

Tato hodnota musí být vyšší než nejnižší hodnota v grafu a zároveň vyšší než hodnota 'y-min', pokud je nastavena.

## X-min

Nastavení 'x-min' slouží k určení minimální zobrazované hodnoty grafu na ose X. 'X-min' může být nastaveno pomocí přepínače '-x', nebo pomocí direktivy 'Xmin' v konfiguračním souboru, a to pouze jednou. Přípustné hodnoty jsou 'min', nebo konkrétní hodnota. Toto nastavení je popsáno v následujících příkladech. Výchozí nastavení je 'min'.

**Příklad nastavení - 'min':** `./circles_graph.py -x min data`

**Konfigurační soubor:** `Xmin min`

**Příklad nastavení - 'hodnota':** `./circles_graph.py -x '10:00:00' -t`

`'[%H:%M:%S]' data`

**Konfigurační soubor:** `TimeFormat [%H:%M:%S]`

`Xmin 10:00:00`

Při nastavení 'x-min' na konkrétní hodnotu bude levý okraj grafu nastaven na tuto hodnotu. Ta nesmí přesahovat všechny vstupní soubory. To by znamenalo že graf bude prázdný a to by vedlo k chybě a k ukončení programu.

## X-max

Nastavení 'x-max' slouží k určení maximální zobrazované hodnoty grafu na ose X. 'X-max' může být nastaveno pomocí přepínače '-X', nebo pomocí direktivy 'Xmax' v konfiguračním souboru, a to pouze jednou. Přípustné hodnoty jsou 'max', nebo konkrétní hodnota. Toto nastavení je popsáno v následujících příkladech. Výchozí nastavení je 'max'.

**Příklad nastavení - 'max':** `./circles_graph.py -X max data`

**Konfigurační soubor:** `Xmax max`

**Příklad nastavení - 'hodnota':** `./circles_graph.py -X '10:00:00' -t`

`'[%H:%M:%S]' data`

**Konfigurační soubor:** `TimeFormat [%H:%M:%S]`

`Xmax 10:00:00`



Při nastavení 'x-max' na konkrétní hodnotu bude pravý okraj grafu nastaven na tuto hodnotu. Ta nesmí být nižší než začátek toho souboru, jehož první řádek má nejnižší časovou hodnotu. To by znamenalo že graf bude prázdný, a to by vedlo k chybě a k ukončení programu. Zároveň musí být také vyšší než hodnota 'x-min', pokud je nastavena.

## Speed

Nastavení 'speed' určuje rychlost růstu grafu ve výsledné animaci. Výchozí hodnota pro toto nastavení je '1'. Definuje se pomocí přepínače '**-S hodnota**', nebo pomocí direktivy '**Speed hodnota**' v konfiguračním souboru. Přípustná hodnota je kladné číslo typu float. Toto nastavení lze definovat pouze jednou.

Nastavení 'time', 'speed' a 'FPS' nemohou být definována současně.

## FPS

Nastavení 'FPS' určuje frame rate výsledné animace. Výchozí hodnota pro toto nastavení je '25'. Definuje se pomocí přepínače '**-F hodnota**', nebo pomocí direktivy '**FPS hodnota**' v konfiguračním souboru. Přípustná hodnota je kladné číslo typu float. Toto nastavení lze definovat pouze jednou.

Nastavení 'time', 'speed' a 'FPS' nemohou být definována současně.

## Time

Nastavení 'time' určuje, jak dlouhá má být výsledná animace. Nemá žádnou výchozí hodnotu. Definuje se pomocí přepínače '**-T hodnota**', nebo pomocí direktivy '**Time hodnota**' v konfiguračním souboru. Přípustná hodnota je kladné číslo typu float. Toto nastavení lze definovat pouze jednou.

Nastavení 'time', 'speed' a 'FPS' nemohou být definována současně.

## Legend

Nastavení legend určuje popis grafu. Nemá žádnou výchozí hodnotu a lze ho definovat pouze jednou. Definuje se pomocí přepínače '**-l hodnota**', nebo pomocí direktivy '**Legend "hodnota"**' v konfiguračním souboru. Přípustná hodnota je libovolný neprázdný řetězec.

## Gnuplot params

Nastavení gnuplot params určuje, jaké parametry se mají předat programu gnuplot při vytváření každého obrázku animace. Toto nastavení nemá žádnou výchozí hodnotu a lze ho použít vícekrát. Přípustná hodnota je libovolný řetězec začínající 'set', nebo 'unset'. Při vícenásobném použití tohoto parametru se gnuplotu předají všechna takto definovaná nastavení.

Lze definovat pomocí přepínače '**-g "hodnota"**', nebo pomocí direktivy '**GnuplotParams hodnota**' v konfiguračním souboru.

## Effect params

Nastavení 'effect params' zahrnuje několik nastavení pro samotný efekt. Tato nastavení se zadávají pomocí přepínače **-e** na příkazové řádce, nebo pomocí direktivy **EffectParams** v konfiguračním souboru. Lze použít vícekrát. Seznam nastavení je oddělený znakem ':' a hodnoty se přidělují znakem '='.

### Příklad nastavení effect params:

```
./circles_graph.py -e columns=40:color=red,blue,green -e  
width=0.6:method=top data
```

Příklad demonstruje možnost nastavení více parametrů efektu pomocí jednoho parametru **-e** a také použití více parametrů **-e** v jednom příkazu.

### Delay

Nastavení 'delay' umožňuje nastavit prodlevu mezi zobrazením každého bodu oproti předchozím. Prodleva může být nastavena na libovolné kladné celé číslo. Výchozí nastavení pro tuto hodnotu je **'10'**.

### Columns

'Columns' určuje, na kolik sloupců (bodů) má být rozdělen celý graf. Toto nastavení neplatí pro vstupní soubor, ale pro celý graf, to znamená, že celkový počet pro jeden soubor (nebo pro více spojených souborů) může být nižší než zadaná hodnota, pokud časový rozsah pro takový soubor nesahá přes celý graf. Výchozí nastavení je závislé na vstupních datech, ale je zaručeno že bude nižší nebo rovno hodnotě **'30'**.

### Color

'Color' umožňuje nastavit barvy pro vykreslované sady bodů. Barvy se zadávají jako seznam oddělený čárkou `'color=red,blue,green'`. Pokud je seznam menší než počet vykreslovaných sad sloupců, tak se pro další sady vyberou náhodné barvy ze všech možných. Barvy se na všechny sady aplikují postupně podle časového pořadí sady bodů. Přípustné barvy jsou všechny barvy podporované programem 'gnuplot'. Tyto barvy lze zjistit zadáním příkazu: `'echo "show colornames" | gnuplot'` (první sloupec výpisu). Výchozí nastavení jsou náhodně vybrané barvy ze seznamu z předchozího příkazu kromě barvy 'white'.

### Method

'Method' určuje použitou metodu pro výpočet konečné hodnoty každého bodu. Přípustné hodnoty pro toto nastavení jsou **'top'**, nebo **'average'**. Výchozí hodnota je **'average'**. Při použití metody 'top' bude pozice bodu nastavena na hodnotu, jejíž absolutní hodnota je nejvyšší ze všech hodnot na řádcích odpovídajících časovému rozsahu pro daný bod (sloupec).

Při použití metody 'average' bude pozice bodu vypočtená zprůměrováním všech hodnot odpovídajících danému rozsahu (sloupci).

## Steps

'Steps' umožňuje určit celkový počet kroků pro celou výšku grafu. Pomocí tohoto nastavení lze tedy definovat velikost posunu (skoku) každého bodu v jednom kroku, a to jako poměr celkové velikosti grafu a 'steps'. Přípustná hodnota je libovolné kladné celé číslo. Výchozí nastavení je '50'.

**Příklad nastavení steps:** `./circles_graph.py -y 5 -Y 15 -e steps=10 data`

Celková velikost posunu (skoku) v předchozím příkladu bude **1**. Výškový rozsah je  $10$  ( $y_{max} - y_{min} = 15 - 5 = 10$ ). Celková velikost přírůstku (skoku) se vypočte jako poměr  $10 / 10 = 1$ .

## Config file

'Config file' určuje, zdali se má načíst nějaký konfigurační soubor. Toto nastavení lze definovat pouze na příkazové řádce pomocí přepínače '-f'. Nemá žádnou výchozí hodnotu. Není možné definovat více konfiguračních souborů.

## Name

'Name' určuje jméno vytvořené animace a adresáře, ve kterém bude uložena. Pokud takový adresář již existuje, připojí se za jméno adresáře sufix '\_i', kde  $i = \max(0, i) + 1$ . Animace má příponu 'mp4'. Výchozí nastavení pro tuto hodnotu je 'circles\_graph'.

Lze definovat pomocí přepínače '-n' na příkazové řádce, nebo pomocí direktivy 'name' v konfiguračním souboru.

**Příklad nastavení name:** `./circles_graph.py -n „graf“ data`

**Konfigurační soubor:** `name graf`

Při nastavení 'name' jako v příkladu bude pro výslednou animaci vytvořen adresář 'graf'.

Pokud takový adresář již existuje, tak jméno nově vytvořeného adresáře bude 'graf\_i', kde 'i' bude nejnižší volná hodnota (viz výše). V tomto adresáři bude uložena samotná animace s názvem 'graf.mp4'. Název animace již nebude obsahovat sufix '\_i'.

'Name' může být libovolný neprázdný řetězec. Toto nastavení lze použít pouze jednou.

## Ignore errors

'Ignore errors' definuje reakci na méně závažné chyby, které nevedou k neschopnosti pokračovat ve vykonávání programu. Při nastavení 'ignore errors' se tyto chyby zobrazují pouze jako varování a program pokračuje dále v činnosti. V případě více závažné chyby, která znemožní další vykonávání programu, bude program ukončen bez ohledu na toto nastavení.

'Ignore errors' lze nastavit z příkazové řádky pomocí přepínače '-E', nebo pomocí direktivy 'IgnoreErrors' v konfiguračním souboru. Ve výchozím nastavení se chyby neignorují.