

# Коротка інструкція щодо користування програмами пакету phonetic-languages-simplified-lists-examples

Автор та розробник ПЗ: Олександр Сергійович Жабенко

Ліцензія: MIT

Copyright (c) 2020 Oleksandr Zhabenko

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

# Зміст

Автор та розробник ПЗ: Олександр Сергійович Жабенко

Ліцензія: MIT

Copyright (c) 2020 Oleksandr Zhabenko

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

Зміст

Вступ

Передумови користування пакетом програм

Встановлення пакету

Робота з програмою lineVariantsG2

Більш комплексне використання

Робота з програмою propertiesTextG2 (та distributionTextG)

I варіант (лише рядки)

II варіант -- статистика по всьому тексту (+ можливо, рядкова)

Копію

[Тип даних](#)

[Значення](#)

[Контроль кількості інтервалів](#)

[Контроль друку також рядка тексту](#)

[Контроль розбивки тексту на рядки](#)

[Той же аргумент щодо кількості рядків](#)

[Чи виводити на екран також і порядкові дані](#)

[Вибірковий аналіз тексту за рядками](#)

[Робота з програмою rewritePoemG2](#)

[Більш комплексне використання](#)

[Типи метрик](#)

[Перший аргумент](#)

[Обмеження \(constraints\)](#)

[Тип обмеження](#)

[Цифрові символи](#)

[Значення обмеження та його параметрів](#)

[Паралельне виконання програм](#)

[Джерела](#)

[Додаткова література](#)

# Вступ

Існують різні мови. У них є різна структура та правила. При цьому існує можливість створити та використовувати (на основі однієї з існуючих широко вживаних та добре розповсюджених мов, зокрема української у цій роботі) “фонетичну” мову, яка краще підходить для поезії та музики. Можливо навіть створити різні варіанти фонетичної мови. Ця робота пропонує створити кілька різних фонетичних мов на основі української.

Уявіть, що ви можете розуміти інформацію в тексті незалежно від порядку слів і при збереженні лише найбільш необхідної граматики (наприклад, правило не відокремлювати прийменник та наступне слово збережене). Розуміти точно так само, як читати текст (після деякого навчання та тренування, можливо), в якому у словах збережені на своїх позиціях лише перші та останні літери, а всі решта -- взаємно перемішані одна з одною. Отже, уявіть, що ви можете розуміти (і виражати ваші думки, почуття, мотиви тощо) повідомлення тексту без дотримання строгого порядку слів.

У такому випадку ви можете впорядкувати слова (зберігши найбільш необхідну граматику для зменшення чи повного усунення можливої двозначності, зумовленої граматикою, точніше зменшенням її обсягу), розмістивши їх таким чином, щоб вони забезпечили собою більш цікаве фонетичне звучання. Ви можете спробувати створити поетичний (чи принаймні дещо більш ритмічний та виразний) текст чи музику. Це також може бути саме по собі розвивальною справою, яка надихає. Але як би ви могли швидко знайти, які комбінації більш чи менш підходять? Крім того, чи може складність алгоритмів бути зменшена?

Це лише деякі з цікавих питань. Ця праця на даний момент не дає повної відповіді на них, але є експериментальною та дослідницькою, при цьому, звичайно, будь-який результат її є цінним.

Українська є мовою без строгих вимог до порядку слів у реченні (хоча є певні усталені переважні варіанти) і має приємне звучання. Отже, вона може бути гарним прикладом та зразком. Крім того, для автора програм це рідна мова.

Навіть якщо ви не бажаєте створити та використовувати “фонетичні” мови, де фонетика є більш важливою, ніж граматика, і тоді ви можете оцінити фонетичний потенціал слів, використаних у тексті, для продукування спеціальним чином озвучених текстів. Це також може бути цінним та помічним у написанні поезії та можливих інших пов’язаних областях.[\[2\]](#)

## Передумови користування пакетом програм

Поки програми працюють для робочих станцій (десктопів, desktop, working station тощо), і немає мобільних версій.

Потрібно, щоб були встановлені та налаштовані програми мовою Haskell:

1. [GHC](#) (версії не раніше 7.10)
2. [Cabal](#)

Виконувані файли цих програм мають бути доступні для пошуку через змінну робочого середовища PATH (це типова їх установка).

## Встановлення пакету

Відкрийте командний рядок чи термінал і введіть як команди:

```
cabal update
```

```
cabal install phonetic-languages-simplified-lists-examples r-glpk-phonetic-languages-ukrainian-durations
```

```
cabal install mmsynbukr (ця остання є опціональною, але корисною для озвучування і не займає багато місця)
```

Якщо є повідомлення про застарілі варіанти команд, введіть замість update -- v1-update, install -- v1-install.

## Робота з програмою **lineVariantsG2**

Перевірте, щоб папка (каталог), куди cabal встановив виконувані файли програм, була доступна для пошуку в змінній середовища PATH.

Далі введіть у командному рядку (чи терміналі) команду:

lineVariantsG2 <[перший аргумент](#)> <[числові аргументи](#)> <[тип метрики](#)> <український текст>

або:

lineVariantsG2 <[перший аргумент](#)> <[числові аргументи](#)> <[тип метрики](#)> <український текст> <[десь серед аргументів як єдина група: обмеження](#)>

та натисніть Enter.

Ви побачите щось на зразок наступного:

lineVariantsG2 10.0\_1.2 ууу садок вишневий коло хати хрущі над вишнями гудуть (введений український текст в кінці команди -- уривок з відомого вірша Тараса Григоровича Шевченка; загалом цей рядок -- це введена команда)

садок колохати хрущі гудуть надвишнями вишневий (варіант (загалом може бути кілька таких варіантів, які утворюють одну групу, а також кілька таких груп; усі групи йдуть одна за одною зверху донизу у порядку зменшення кінцевого значення метрики), який максимізує обрану метрику для заданих інтервалів)

[ 5.5438 ] (значення обраної метрики до застосування перетворення інтервалів)

[ 5.5438 ] (значення обраної метрики після перетворення інтервалів, кінцеве значення метрики для цього рядка)

Зверніть увагу, що текст може (і переважно буде) писатися не так, як він пишеться згідно правил орфографії та пунктуації, але ви можете його прочитати та спробувати зрозуміти. Змінюючи введені перші аргументи, ви (скоріше всього) отримуватимете інші вихідні дані, те ж, вочевидь, стосується й українського тексту. Занадто довгий текст буде скорочений до об'єму, який ви змогли б зрозуміти (можливо, після згаданого раніше тренування) без надзусиль.

Спробуйте оцінити, прочитавши варіант, наскільки він підходить.

## Більш комплексне використання

Числові аргументи, якщо задані, мають наступне значення.

Перший числовий аргумент -- кількість груп з однаковими максимальними значенням метрики (в порядку зменшення), які будуть виведені на екран як результат. Якщо задано більшу кількість, ніж їх є взагалі, то виводяться всі можливі результати, які задовольняють усі інші умови. Якщо не задано числові аргументи, то вважається рівним 1.

Другий числовий аргумент -- кількість інтервалів, на які поділяється проміжок між мінімальним та максимальним значенням метрики для даного рядка. Якщо не задано, вважається рівним 1. Значення 0 не дає змоги іншим числовим аргументам далі змінювати результат роботи програми.

Усі наступні числові аргументи (якщо задано, інакше ніяких перестановок не відбувається) -- номери інтервалів, які будуть поміняні місцями з максимальним за номером. Це дозволяє змінити структуру даних, які відображаються як результат роботи програми і побачити внутрішні (не максимальні) елементи. Наприклад, числові аргументи 2 6 1 4 (у такому порядку) означатимуть, що в ході виконання програма поверне 2 групи елементів з максимальними значеннями метрики (найбільшим і найбільшим наступним за попереднім), отриманим після перестановки інтервалів; відрізок між максимальним та мінімальним значенням метрики буде поділено на 6 рівних інтервалів, при цьому елементи, які знаходяться у першому та 4, рахуючи від мінімального (інтервал з номером 1) виведе максимальні 2 групи елементів.

Значення, які були в максимальному інтервалі, будуть переміщені в інтервал з найменшим номером серед тих, які переміщені в максимальний. Таким чином, при виводі ці значення будуть виведені найпізніше.

Після кожного рядка буде відображено 2 числа у квадратних дужках -- первинне значення метрики (без переміщення інтервалів) і значення після переміщення.

Про використання інших параметрів трохи згодом.

## Робота з програмою **propertiesTextG2** (та **distributionTextG**)

### I варіант (лише рядки)

Перевірте, щоб папка (каталог), куди cabal встановив виконувані файли програм, була доступна для пошуку в змінній середовища PATH.

Далі введіть у командному рядку (чи терміналі) команду:

```
propertiesTextG2 <перший аргумент> <файл з українським текстом> <контроль кількості інтервалів> <контроль друку також рядка тексту> <контроль розбивки тексту на рядки> <тип метрики>
```

та натисніть Enter.

Ви побачите щось схоже на:

```
propertiesTextG2 2.1_3.0 ~/sadok.txt s 1 0 02y
```

108	108	142	1.0000	1.3111	1.3111	0.8653846	3	1	Садок вишневий колохати
193	200	200	1.0387	1.0387	1.0000	1.0189831	3	4	Хрущі надвишнями гудуть
192	200	200	1.0437	1.0437	1.0000	1.0214068	3	4	Плугатарі зплугами йдуть
108	109	110	1.0049	1.0194	1.0145	0.9951857	3	2	Співають ідучи дівчата
192	193	200	1.0049	1.0437	1.0387	0.98335034	3	1	Аматері вечерять ждуть
108	245	245	2.2694	2.2694	1.0000	1.3882753	3	4	Сем'я вечеря колохати
192	200	200	1.0437	1.0437	1.0000	1.0214068	3	4	Вечірня зіронька встає
192	200	226	1.0437	1.1750	1.1257	0.9597701	3	1	Дочка вечерять подає
108	277	277	2.5611	2.5611	1.0000	1.4383775	3	4	Амати хоче научати
192	192	193	1.0000	1.0049	1.0049	0.9975753	3	1	Так соловейко недає
108	142	245	1.3111	2.2694	1.7309	0.8020391	3	1	Поклала мати колохати
192	192	200	1.0000	1.0437	1.0437	0.9785933	3	1	Маленьких діточок своїх
192	192	200	1.0000	1.0437	1.0437	0.9785933	3	1	Сама заснула колойіх
27	29	242	1.0778	8.9778	8.3299	0.21603565	4	1	Затихло все тільки дівчата
192	193	193	1.0049	1.0049	1.0000	1.0024246	2	4	Тасоловейко незатих

Числові стовпчики мають однакове значення для обох варіантів для рядків. Відмінність полягає у тому, що у другому випадку статистика по всьому тексту має більше значення з погляду дослідника (дослідниці), ніж по кожному рядку зокрема.

I стовпчик -- це мінімально можливе значення обраної метрики для заданих даних серед усіх можливих варіантів перестановок слів у рядку;

II стовпчик -- це актуальне значення обраної метрики для заданих даних у рядку, те, яке реалізується у конкретно цьому варіанті рядка;



III стовпчик -- це максимально можливе значення обраної метрики для заданих даних серед усіх можливих варіантів перестановок слів у рядку;

IV стовпчик -- це відношення значення метрики для даного рядка та її мінімального значення для слів, з яких складається рядок; число, яке не менше 1.0;

V стовпчик -- це відношення максимального значення метрики для слів даного рядка та її мінімального значення, з яких складається рядок; число, яке не менше 1.0 і не менше числа у IV стовпчику;

VI стовпчик -- це відношення максимального значення метрики для даного рядка та її актуального значення; число, яке не менше 1.0 і не більше значення у V стовпчику;

VII стовпчик -- це відношення актуального значення метрики до середнього арифметичного (півсуми) максимального та мінімального значення для всіх можливих перестановок слів при заданих даних; число, яке відображається з повною обчисленою кількістю знаків після крапки; має важливе значення для подальшої статистики для всього тексту;

VIII стовпчик -- кількість слів у рядку, деякі з яких можуть складатися з кількох з'єднаних українських слів;

IX стовпчик -- номер інтервалу (починаючи з 1), до якого входить актуальне значення метрики для заданих даних;

Далі направо -- якщо вказано <контроль друку також рядка тексту> як "1", тоді тут виводиться рядок тексту, який аналізується; інакше, ці дані не виводяться.

## II варіант -- статистика по всьому тексту (+ можливо, рядкова)

Перевірте, щоб папка (каталог), куди cabal встановив виконувані файли програм, була доступна для пошуку у змінній середовища PATH.

Далі введіть у терміналі команду:

```
propertiesTextG2 <перший аргумент> <файл з українським текстом> <контроль кількості інтервалів> <контроль  
друку також рядка тексту> <контроль розбивки тексту на рядки> <тип метрики> | distributionTexG <той же  
аргумент щодо кількості рядків> <чи виводити на екран також порядкові дані>
```

та натисніть Enter. У Unix-подібних ОС вертикальна лінія, виділена червоним, слугує для створення пайплайнів (pipelines) у терміналі в shell; для ОС Windows :

```
PowerShell -Command "propertiesTextG2 <перший аргумент> <файл з українським текстом> <контроль кількості інтервалів> <контроль друку також рядка тексту> <контроль розбивки тексту на рядки> <тип метрики> | distributionTexG <той же аргумент щодо кількості рядків> <чи виводити на екран також порядкові дані>"
```

Ви побачите щось схоже на:

```
propertiesTextG2 2.1_3.0 ~/sadok.txt s 1 0 02y | distributionTextG s 1
```

108	108	142	1.0000	1.3111	1.3111	0.8653846	3	1	Садок вишневий колохати
193	200	200	1.0387	1.0387	1.0000	1.0189831	3	4	Хрущі надвишнями гудуть
192	200	200	1.0437	1.0437	1.0000	1.0214068	3	4	Плугатарі зплугами йдуть
108	109	110	1.0049	1.0194	1.0145	0.9951857	3	2	Співають ідучи дівчата
192	193	200	1.0049	1.0437	1.0387	0.98335034	3	1	Аматері вечерять ждуть
108	245	245	2.2694	2.2694	1.0000	1.3882753	3	4	Сем'я вечеря колохати
192	200	200	1.0437	1.0437	1.0000	1.0214068	3	4	Вечірня зіронька встає
192	200	226	1.0437	1.1750	1.1257	0.9597701	3	1	Дочка вечерять подає
108	277	277	2.5611	2.5611	1.0000	1.4383775	3	4	Амати хоче научати
192	192	193	1.0000	1.0049	1.0049	0.9975753	3	1	Так соловейко недає
108	142	245	1.3111	2.2694	1.7309	0.8020391	3	1	Поклала мати колохати
192	192	200	1.0000	1.0437	1.0437	0.9785933	3	1	Маленьких діточок своїх
192	192	200	1.0000	1.0437	1.0437	0.9785933	3	1	Сама заснула колойіх
27	29	242	1.0778	8.9778	8.3299	0.21603565	4	1	Затихло все тільки дівчата
192	193	193	1.0049	1.0049	1.0000	1.0024246	2	4	Тасоловейко незатих

1 2 3 4

8 1 0 6

53.33% 6.67% 0.00% 40.00%

0.9778+-0.2597 0 15

2 3 4 5 6 7

```

1      13      1      0      0      0
*****
.      .      .      1
7      1      .      5
1      .      .      .
.      .      .      .
.      .      .      .
.      .      .      .
~~~~~
0      0      0      1
7      1      0      5
1      0      0      0
0      0      0      0
0      0      0      0
0      0      0      0
=====

```

Виділення кольором не буде, тут просто згруповані різні типи даних статистики.

Колір	Тип даних	Значення
Червоний	Десятковий дріб з похибкою	<p>Середнє арифметичне значення (математичне сподівання) всіх чисел у VII стовпчику статистики для рядків, плюс-мінус середнє квадратичне відхилення; у випадку обраної метрики “y0” -- часто число близьке до 1.0; може містити похибку округлення викликану особливостями знаходження суми чисел з плаваючою крапкою.</p> <p>У випадку, якщо всі рядки тексту виключаються з аналізу (дивіться: пояснення для</p>

		<p>оранжевого кольору), то виводиться замість матриць та подальшої інформації лише один рядок тексту -- сигнальний напис:</p> <p>"1.000+-0.000\tALL!",</p> <p>де '\t' означає символ табуляції.</p> <p>Цей напис означає, що вказаний текст не підходить для аналізу програмою, а також, що будь-які дані, які в ході роботи програми можна було б отримати з цього тексту та супутніх текстів, повинні аналізуватися так, щоб не спростувати зроблені висновки на основі усієї сукупності текстів. Простіше кажучи, у такому випадку можна не враховувати текст, бо при правильному підході до аналізу та інтерпретації даних він не повинен скасувати результати.</p>
Синій	Натуральні числа	<p>Матриця розподілу рядків за кількістю слів та інтервалами; елемент матриці у k-му рядку та j-му стовпчику -- кількість рядків, для яких значення обраної метрики з вказаними даними потрапляє в інтервал з номером j (нумерація починається з 1), який рівний номеру стовпчика матриці, і містить у рядку k слів (слів або їх сполучень, які відображаються злитно для дотримання мінімальних правил граматики при аналізі та виведенні даних), число k лежить у діапазоні між 2 і 7 включно (рядки, які аналізуються для матриці, містять від 2 до 7 слів (або написаних злитно сполучень)). Таким чином, матриця завжди має 6 рядків, а число стовпчиків буде залежати від введених та наявних даних.</p> <p>Фактично ця матриця заміняє графік двовимірного розподілу даних.</p> <p>На екран виводиться двічі, одна від іншої відмежована тильдами. У першому випадку нульові значення не відображаються, замість них стоять крапки. Це елемент візуалізації даних, який дозволяє краще "побачити", як виглядає розподіл, де числові значення відповідають "висоті" на графіку розподілу (значенню функції дискретного двовимірного розподілу). У другому випадку на місці крапок стоять відповідні значення, які всі рівні 0. Дані для матриці отримуються з VIII та IX стовпчиків статистики за рядками.</p>
Оранжевий	Цілі невід'ємні числа	<p>Загальні кількості рядків. Перше число зліва -- кількість рядків, які виключаються з аналізу для матриці, оскільки в них мало даних (1 або менше слів). Рівність 0 означає, що усі рядки виведені на екран беруть участь в аналізі для утворення матриці розподілу. Число справа -- загальна кількість рядків у тексті, який виводиться та аналізується (включно з тими рядками, які пораховані зліва).</p>
Зелений	Відсотки	<p>Розподіл загальної кількості рядків за інтервалами. Сума відсоткових значень природно</p>

		<p>рівна 100%. Номери інтервалів надписані зверху над відповідними відсотковими значеннями. Наприклад, напис у цих трьох рядках типу:</p> <p>1     2 10   15 40% 60%</p> <p>означає, що з усієї загальної кількості рядків, які піддаються аналізу за допомогою програми (містять достатньо даних), 40% припадає на перший інтервал (з меншим значенням метрики), а 60% -- на другий (відповідно з більшим значенням метрики). Тобто таких рядків відповідно 10 і 15.</p> <p>Усі інтервали рівні за величиною, але можуть мати різні кількості рядків (яка знаходиться в ході роботи програми).</p> <p>Одновимірний розподіл, за ним можна побудувати гістограму.</p>
Бузковий	Натуральні числа	Номери інтервалів. Починається відлік з 1. Під ними відповідні кількості рядків, значення метрики для яких згідно даних потрапляє у відповідний за номером інтервал.
Фіолетовий	Натуральні числа	Кількість слів у рядках. Лежить в межах від 2 до 7 включно (якщо слів менше, тоді рядок отримує значення рівні 1.0 і вилучається з аналізу програмою для матриці). Під ними -- відповідні значення кількості таких рядків. 0 відповідає випадку відсутності рядків з даною кількістю слів (чи сполучень, які відображаються як одне слово).

Уважне вивчення цих даних дає змогу зробити певні висновки щодо тексту, їх сукупності, самої моделі та мови.

## Контроль кількості інтервалів

Можливі три випадки:

“**s**” -- кількість інтервалів буде визначена за відомим [правилом Стерджеса](#), де кількість випробувань буде рівною результуючій кількості рядків;

“**I**” -- кількість інтервалів буде визначена за рекомендацією В. П. Левинського (див.: Опря А. Т. Статистика (модульний варіант з програмованою формою контролю знань). -- Навч. посіб. --- К.: Центр учбової літератури, 2012. --- 448 с. ISBN 978-611-01-0266-7. С. 60);

**число** -- кількість інтервалів буде задане натуральне число (має бути більше 1, хоча це не перевіряється);

**щось інше** -- буде використано 9.

## Контроль друку також рядка тексту

Якщо цей аргумент рівний 1, то праворуч від числових даних порядкової статистики буде виведено на екран також рядок, який аналізується (в уже перетвореному вигляді для аналізу). Інакше рядок не буде виведений.

## Контроль розбивки тексту на рядки

Якщо задати тут 1, то текст буде спочатку згрупований в один рядок, а потім розбитий на рядки методом ділення навпіл (за кількістю слів чи їх сполучень) доти, поки довжина усіх рядків не буде менша за 8 слів чи їх сполучень. Сторонні символи при цьому будуть відфільтровані. Якщо задати 0, то текст буде аналізуватися після фільтрації сторонніх символів (приблизно) у тих рядках, які були спочатку.

## Той же аргумент щодо кількості рядків

Мається на увазі тут має стояти те ж [значення](#), що і на місці контролю кількості інтервалів.

## Чи виводити на екран також і порядкові дані

Тут потрібно поставити 1, щоб програма надрукувала всю статистику (спочатку порядкову, а потім загальну по тексту), інакше буде надруковано лише загальну по тексту.

## Вибірковий аналіз тексту за рядками

Якщо Ви виконаєте команду

```
propertiesTextG2 <шлях до файлу з українським текстом для аналізу> @n
```

то на екран буде виведений текст з файлу, який буде аналізований з номерами усіх рядків зліва від самих рядків, що відділені від тексту символом табуляції (відображається у вигляді пробілу з непостійною шириною, яка залежить від налаштувань системи).

Тоді Ви можете виконувати цю ж програму (можна і без того, але можете вказати інші номери, ніж вважатиме програма) для аналізу вибраних рядків. Для цього до команд `propertiesTextG2`, крім останньої згаданої, будь-де у рядку команди до вертикальної риски (до пайплайну) додайте номери першого та останнього рядка, відділені символом двокрапки (без будь-яких інших символів, зокрема без пробілів). Можна задавати декілька таких пар, інформація буде виведена у такому ж порядку. Якщо деякі номери рядків будуть зустрічатися кілька разів, вони будуть виведені (якщо вказана така опція) та проаналізовані так само кілька разів. Якщо задано контроль розбивки тексту на рядки, що рівний 1, тоді програма об'єднає і проаналізує ті рядки, номери яких були вказані і відповідають номерам при виводі команди з @n.

Це все дозволяє сфокусовано аналізувати текст чи лише його частини.

## Робота з програмою **rewritePoemG2**

Перевірте, щоб папка (каталог), куди cabal встановив виконувані файли програм, була доступна для пошуку в змінній середовища PATH.

Далі введіть у командному рядку (чи терміналі) команду:

```
rewritePoemG2 <перший аргумент> <файл з українським текстом> <тип метрики> <числові аргументи>  
та натисніть Enter.
```

Ви побачите щось на зразок наступного:

```
rewritePoemG2 10.0_1.2 "sadok.txt" yyy 5 1 2
```

Після успішного завершення виконання програми (не має бути жодних повідомлень) тій же папці (каталозі), де і файл з текстом, що переписується, має бути файл з додатковим закінченням `.new.txt`. Саме там записаний перетворений текст (наприклад, вірш) згідно введених даних.

Введені дані стосуються усього тексту, тобто кожного рядка тексту зокрема (після його попередньої обробки програмою).

## Більш комплексне використання

Числові аргументи мають багато в чому схоже значення, як і для програми [lineVariantsG2](#).

Числові аргументи, якщо задані, мають наступне значення.

Перший числовий аргумент -- кількість інтервалів, на які поділяється проміжок між мінімальним та максимальним значенням метрики для даного рядка. Якщо не задано, вважається рівним 1. Значення 0 не дає змоги іншим числовим аргументам далі змінювати результат роботи програми.

Усі наступні числові аргументи (якщо задано, інакше ніяких перестановок не відбувається) -- номери інтервалів, які будуть поміняні місцями з максимальним за номером. Це дозволяє змінити структуру даних, які відображаються як результат роботи програми і побачити внутрішні (не максимальні) елементи. Наприклад, числові аргументи 6 1 4 (у такому порядку) означатимуть, що в ході виконання програми відрізок між максимальним та мінімальним значенням метрики буде поділено на 6 рівних інтервалів, при цьому елементи, які знаходяться у першому та 4, рахуючи від мінімального (інтервал з номером 1) буде переміщено до максимального за номером (і значеннями метрики) інтервалу, а потім записано у файл виводу результатів рядок з максимальним значенням метрики.

Значення, які були в максимальному інтервалі, будуть переміщені в інтервал з найменшим номером серед тих, які переміщені в максимальний.

## Типи метрик

Одним з принципів роботи програми є пошук серед варіантів тексту тих, для яких максимальним є значення певної функції, яка називається "властивістю" тексту (property) і становить собою певну метрику для рядків. Користувач може самостійно обрати метрику, яка буде використана при роботі програми (це робиться в командному рядку один раз протягом роботи програми заданням (або відсутністю відповідного) командного аргумента).

Аргумент командного рядка виклику програми (для версії `phonetic-languages-examples-0.3.1.1`, для новіших версій список очікувано може бути більшим) може бути:

- **"y0"** -- найперша у часі метрика, заснована на ["періодах унікальності"](#). Ідея полягає у тому, що оцінюються кількості звуків, або пауз, або фонетичних явищ (палаталізації приголосних), які знаходяться між послідовними появами кожного звука не в одному, а в різних словах, і шукається загальна сума таких відстаней для різних слів. Більшому значенню відповідає текст з більш плавно змінним фонетичним малюнком, (імовірно) у залежності від середньої кількості звуків у "періоді унікальності" його може бути легше чи важче промовляти; меншому значенню -- навпаки -- текст з більш стрімкими змінами фонетичного малюнку, можливо, з підсиленням окремих груп звуків, що більш характерно для інтонаційно виділених і / чи поетичних текстів з



закликами чи підсиленими емоціями. При використанні цієї метрики (і виключно її) перший аргумент рядка виклику програми не має значення (він ігнорується програмою).

- “0y” -- перша версія метрики аналізу лише ритмічності. Метрика (напівемпірична), заснована на функції ритмічності, що використовує тривалості звуків, які були синтезовані у пакеті програм [mmsyn6ukr](#). Функція ритмічності натхненна [античною поезією](#), де замість наголошених та ненаголошених складів чергувалися ритмічно короткі та довгі; також музичними долями, для яких основними є дводольний ритм та тридольний. Функція реалізована таким чином, щоб можливо простіше вловити значні викиди підритмічностей для двоскладових і трискладових випадків. Використовуючи <перший аргумент> можна змінити співвідношення цих підметрик і відповідно -- змінити метрику.
- “02y” -- подібна до “0y” метрика, яка використовує інші тривалості звуків, синтезовані завдяки пакету [r-glphk-phonetic-languages-ukrainian-durations](#). Можливо, найточніша з цих наявних для задачі написання ритмічного тексту. Можна створити інші варіанти тривалостей звуків, використовуючи можливості пакету r-glphk-phonetic-languages-ukrainian-durations або іншим способом.
- “y” -- метрика, яка обчислює метрики “y0” та “0y” у більш ефективний спосіб, ніж кожен з них поодиночі, а потім перемножує отримані дані. Дає більші значення для рядків з більш плавно змінним фонетичним малюнком та такі, які більше ритмізовані (з точки зору метрики “0y”). Використання <першого коефіцієнта> внутрішньо впливає лише на підметрику “0y”.
- “y2” -- метрика, подібна до y, але використовує замість другої підметрики (ритмічності) варіант з “02y”.
- “yy” -- метрика, яка використовує метрику “y0” та “0y”, при цьому замість їх перемноження, ділить результат другої на результат першої. Максимізується для текстів з високою ритмічністю (з точки зору метрики “0y”) та групуванням однакових звуків у групи ближче один до одного. Використання <першого коефіцієнта> впливає лише на підметрику “0y”.
- будь-які **інші варіанти** даного аргумента -- аналогічна до “yy” з тією відмінністю, що замість “0y” використовується “02y”.

## Перший аргумент

При ознайомленні програми можна використовувати без цього аргументу, або ввівши на його місці 1\_. Надалі Ви, можливо, захочете поглибити аналіз. Тоді можна задати також перший аргумент (він йде першим у списку аргументів командного рядка, не рахуючи групу -- якщо така наявна -- між аргументами +A та -A) у вигляді число1\_число2, де число1 та число2 -- десяткові додатні дробі або цілі додатні числа, причому одне з них може бути відсутнім (тоді воно вважається рівним 1). Наприклад,

3.4\_2

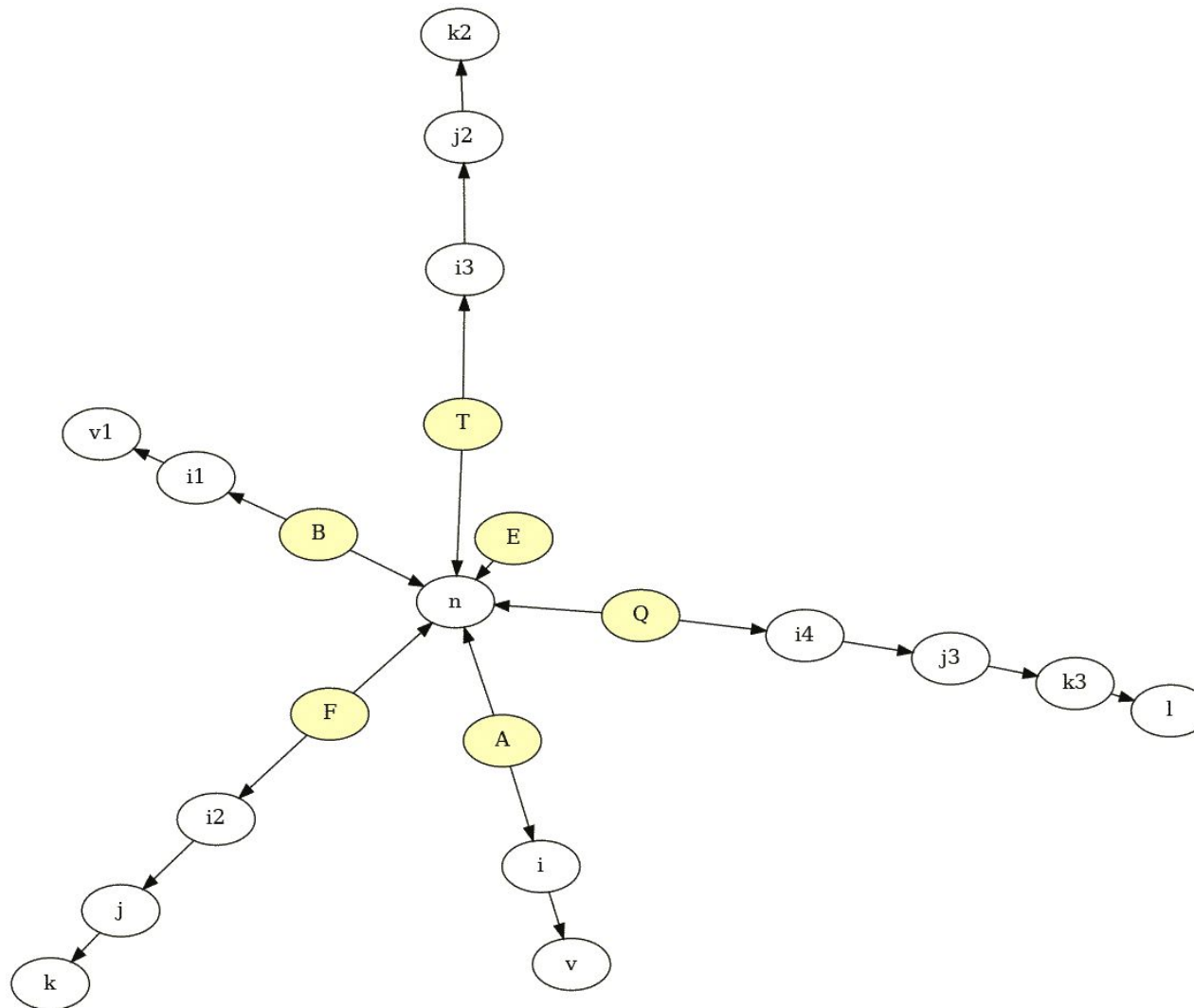
2.987\_0.7865

0.0001\_

тощо.

Тоді перше число буде коефіцієнтом, який множиться на складову метрики, яка відповідає за ритмічність двоскладових стоп, а другий -- за ритмічність трискладових стоп. Відповідно вдалою комбінацією можна підкресли, скомбінувати чи знівелювати вплив ритмічності для двоскладових чи трискладових стоп.

За відсутності цього аргументу програми поведуть себе так, наче він рівний 1\_1.



## Обмеження (constraints)

При запуску програми [lineVariantsG2](#) можна вказати як аргументи командного рядка обмеження. Вони дозволяють зменшити кількість обчислень, розглядати лише певні варіанти (наприклад, з певним визначенням порядком деяких слів тощо), що дозволяє фактично розширити можливості програми. Ці обмеження кодуються як аргументи командного рядка між двома спеціальними позначеннями +A та -A. Вони утворюють групу аргументів, які можуть стояти будь-де у рядку вводу даних. У залежності від того, чи вказуються ці аргументи, програма ставить чи не ставить додаткове питання для перевірки та підтвердження (так званий *double check*).

Типів обмежень є 6, їх можна довільно комбінувати, але з дотриманням меж для кожного з них.

На рисунку поряд видно, що усі типи реалізовані з одним аргументом, який

подібний у всіх них -- це кількість слів (чи їх сполучень) у рядку. Користувач, запустивши програму, уже не може коригувати в ході її роботи цю кількість, але вона важлива для обмежень загалом. Жодний з цифрових символів у обмеженнях не повинен бути більшим за це число, також це число саме не більше 6 і не менше 0. Також необхідною умовою є те, що жодні цифрові символи в межах одного кодованого обмеження не можуть повторюватися двічі. Наприклад, завідомо не є валідними обмеження: Q2235 (повтор цифр), E2 (цифрові символи там, де їх немає), T247 (7 більша 6), F0 (один символ замість необхідних двох), A37523 (7 більша 6), B5 (один символ, а мають бути ще). Неправильно задані обмеження або не вплинуть на результат (хоча буде очікуватися інше), або викличуть помилку виконання (runtime exception) і зупинку роботи програми. Оскільки результат їх застосування не є простим, тому програма при їх заданні виводить на екран рядок, до якого будуть застосовані введені обмеження з додатковим запитанням, чи всі дані введено правильно.

Типи обмежень та їх значення наведено детальніше у таблиці.

Тип обмеження	Цифрові символи	Значення обмеження та його параметрів
<b>E</b>	Без вводу додаткових цифрових символів.	Відповідає відсутності додаткового обмеження, в ході роботи програми не впливає на кінцевий результат.
<b>Q</b>	4 попарно нерівні цифри в межах від 0 до кількості слів чи їх сполучень мінус 1	Цифри -- це індекси 4-х слів чи їх сполучень, взаємний порядок яких при перестановках буде збережений таким. Також якщо ці слова однакові (без врахування великих та малих літер), тоді це зручний спосіб зменшити об'єм даних, який буде аналізуватися.
<b>T</b>	3 попарно нерівні цифри в межах від 0 до кількості слів чи їх сполучень мінус 1	Цифри -- це індекси 3-х слів чи їх сполучень, взаємний порядок яких при перестановках буде збережений таким. Також якщо ці слова однакові (без врахування великих та малих літер), тоді це зручний спосіб зменшити об'єм даних, який буде аналізуватися.
<b>F</b>	2 попарно нерівні цифри в межах від 0 до кількості слів чи їх сполучень мінус 1	Цифри -- це індекси 2-х слів чи їх сполучень, взаємний порядок яких при перестановках буде збережений таким. Також якщо ці слова однакові (без врахування великих та малих літер), тоді це зручний спосіб зменшити об'єм даних, який буде аналізуватися.

<b>A</b>	1 цифра та ще кілька попарно нерівних цифр (усі між собою нерівні) справа від неї в межах від 0 до кількості слів чи їх сполучень мінус 1	Перша цифра -- індекс елемента, відносно якого визначається розміщення усіх інших елементів (слів чи їх сполучень); усі інші цифри правіше -- індекси елементів, які мають стояти в отриманих перестановках ПРАВИШЕ від елемента з індексом, рівним першій цифрі.
<b>B</b>	1 цифра та ще кілька попарно нерівних цифр (усі між собою нерівні) справа від неї в межах від 0 до кількості слів чи їх сполучень мінус 1	Перша цифра -- індекс елемента, відносно якого визначається розміщення усіх інших елементів (слів чи їх сполучень); усі інші цифри правіше -- індекси елементів, які мають стояти в отриманих перестановках ЛІВІШЕ від елемента з індексом, рівним першій цифрі.

## Паралельне виконання програм

Зазвичай усі програми пакету виконуються одним ядром процесора. При цьому для всіх програм, які розглядаються, є можливість включити режим роботи на кількох ядрах -- паралельні обчислення. Для цього серед аргументів командного рядка мають бути наступні:

+RTS -N -RTS

Їх розміщення не впливає на порядок та значення інших аргументів командного рядка, також між записами RTS можуть бути інші параметри. Детальніше про ці параметри дивіться документацію англійською [3].

## Джерела

1. Шевченко Т. Г. "Садок вишневий коло хати...". [Електронний ресурс]. Режим доступу: [Sadok vyshnevyi kolo khaty](#). Перевірено 09 листопада 2020 р.
2. Oleksandr Zhabenko. dobutokO-poetry. [Електронний ресурс]. Режим доступу: [dobutokO-poetry: Helps to order the 7 or less Ukrainian words to obtain somewhat suitable for poetry or music text](#). Перевірено 09 листопада 2020 р.

3. Glasgow Haskell Compiler User's Guide. 7.5. Using SMP parallelism. [Електронний ресурс]. Режим доступу: [https://downloads.haskell.org/~ghc/latest/docs/html/users\\_guide/using-concurrent.html#using-smp-parallelism](https://downloads.haskell.org/~ghc/latest/docs/html/users_guide/using-concurrent.html#using-smp-parallelism) . Перевірено 10 листопада 2020 р.
4. Ямбічна строфа // [Енциклопедичний словник класичних мов](#) / [Л. Л. Звонська](#), Н. В. Корольова, О. В. Лазер-Паньків та ін.. — К. : ВПЦ «Київський університет», 2017. — С. 551. — 552 с.
5. Смаглій Г. А. Теорія музики : Підруч. для навч. закл. освіти, культури і мистецтв / Г. А. Смаглій. — Х. : Вид-во “Ранок”, 2013. — 392 с. ISBN 978-617-09-1294-7

## Додаткова література

6. [Квантитативне віршування](#) — [Вікіпедія](#) [Електронний ресурс]. Режим доступу: [https://uk.wikipedia.org/wiki/%D0%9A%D0%B2%D0%B0%D0%BD%D1%82%D0%B8%D1%82%D0%B0%D1%82%D0%B8%D0%B2%D0%BD%D0%B5\\_%D0%B2%D1%96%D1%80%D1%88%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F](https://uk.wikipedia.org/wiki/%D0%9A%D0%B2%D0%B0%D0%BD%D1%82%D0%B8%D1%82%D0%B0%D1%82%D0%B8%D0%B2%D0%BD%D0%B5_%D0%B2%D1%96%D1%80%D1%88%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F). Перевірено 12 листопада 2020 р.
7. [Силабічне віршування](#) — [Вікіпедія](#) [Електронний ресурс]. Режим доступу: [https://uk.wikipedia.org/wiki/%D0%A1%D0%B8%D0%BB%D0%B0%D0%B1%D1%96%D1%87%D0%BD%D0%B5\\_%D0%B2%D1%96%D1%80%D1%88%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F](https://uk.wikipedia.org/wiki/%D0%A1%D0%B8%D0%BB%D0%B0%D0%B1%D1%96%D1%87%D0%BD%D0%B5_%D0%B2%D1%96%D1%80%D1%88%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F). Перевірено 12 листопада 2020 р.
8. [Тонічне віршування](#) — [Вікіпедія](#) [Електронний ресурс]. Режим доступу: [https://uk.wikipedia.org/wiki/%D0%A2%D0%BE%D0%BD%D1%96%D1%87%D0%BD%D0%B5\\_%D0%B2%D1%96%D1%80%D1%88%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F](https://uk.wikipedia.org/wiki/%D0%A2%D0%BE%D0%BD%D1%96%D1%87%D0%BD%D0%B5_%D0%B2%D1%96%D1%80%D1%88%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F). Перевірено 12 листопада 2020 р.
9. [Силабо-тонічне віршування](#) — [Вікіпедія](#) [Електронний ресурс]. Режим доступу: [https://uk.wikipedia.org/wiki/%D0%A1%D0%B8%D0%BB%D0%B0%D0%B1%D0%BE-%D1%82%D0%BE%D0%BD%D1%96%D1%87%D0%BD%D0%B5\\_%D0%B2%D1%96%D1%80%D1%88%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F](https://uk.wikipedia.org/wiki/%D0%A1%D0%B8%D0%BB%D0%B0%D0%B1%D0%BE-%D1%82%D0%BE%D0%BD%D1%96%D1%87%D0%BD%D0%B5_%D0%B2%D1%96%D1%80%D1%88%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F). Перевірено 12 листопада 2020 р.
10. [Віршування](#) — [Вікіпедія](#) [Електронний ресурс]. Режим доступу: <https://uk.wikipedia.org/wiki/%D0%92%D1%96%D1%80%D1%88%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F>. Перевірено 12 листопада 2020 р.

11. [Античне віршування — Вікіпедія](https://uk.wikipedia.org/wiki/%D0%90%D0%BD%D1%82%D0%B8%D1%87%D0%BD%D0%B5_%D0%B2%D1%96%D1%80%D1%88%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F) [Електронний ресурс]. Режим доступу: [https://uk.wikipedia.org/wiki/%D0%90%D0%BD%D1%82%D0%B8%D1%87%D0%BD%D0%B5\\_%D0%B2%D1%96%D1%80%D1%88%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F](https://uk.wikipedia.org/wiki/%D0%90%D0%BD%D1%82%D0%B8%D1%87%D0%BD%D0%B5_%D0%B2%D1%96%D1%80%D1%88%D1%83%D0%B2%D0%B0%D0%BD%D0%BD%D1%8F). Перевірено 12 листопада 2020 р.

12.