

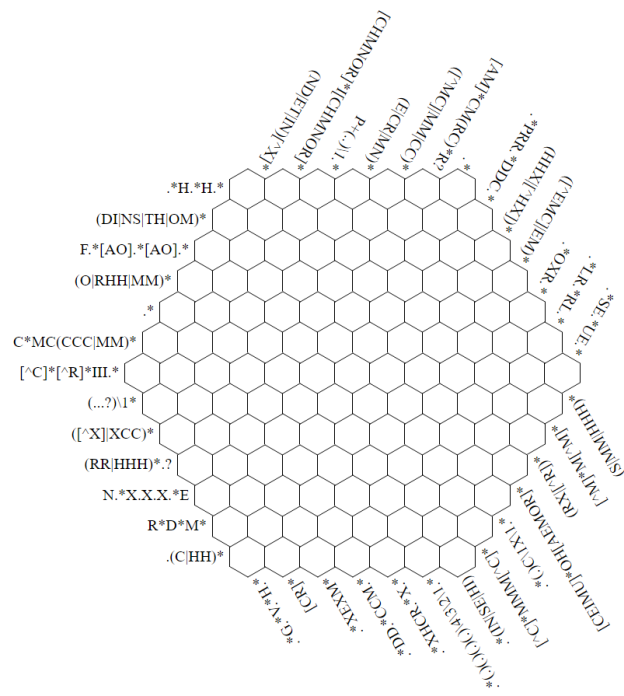


Курсовая работа

Создание программы генерирующей кроссворды из регулярных выражений

Данилевич Леонид
Лельчук Александр
Под руководством Дворкина
Михаила Эдуардовича

Кроссворды из регулярных выражений



Кроссворд из регулярных выражений -- клеточная сетка правильной формы (прямоугольник, шестиугольник и т.д.), который необходимо заполнить латинскими буквами так, что каждая линия (строка или столбец в прямоугольнике) подходит под соответствующее регулярное выражение.



Цель

Написать программу, функционалом которой является генерация, отображение и поддержка решения пользователем красивых и сложных кроссвордов из регулярных выражений.

Решение кроссворда

Решение кроссворда:

1. Парсинг регулярных выражений – разбор на токены, составление “слов”
2. Инкрементальные улучшения – отгадывание букв по очереди в некотором порядке

Бонус: оценка сложности (число попыток взятия клеточки с ещё не отгаданной буквой). Усреднение по 10 попыткам решения, порядок случайный.

(.A?)B\1.*

5 букв



В	А	В	В	А
1	2	3	1	2

В	В	В	В	В
1	2	1	3	4

Как генерируются поля и регулярные выражения?

Центральный алгоритм — метод имитации отжига.

1. Генерация поля кроссворда

Оценка стоимости символьной строки — поиск паттернов:

Повторения (ABABAB); словарные ключи (MASK);
палиндромы (ABCBA), прогрессии (AXAYX).

Как генерируются поля и регулярные выражения?

Центральный алгоритм — метод имитации отжига.

2. Генерация регулярных выражений для кроссворда

Начальное состояние — побуквенные описания.

Оценка стоимости:

Коэффициенты для специальных символов

(* + . \ []).

G R E E T I N G
U A U A U A U A
M R M R A I N W
E V O K E A E A
N R N R N R N R
E A O E E O E E
M R M R A P N J
U E U E R E E L

R C G R E E T I N G
I S S U E P D Q D P T
D H T Z E U U S O B T
S U I T U Q F W G I G
T R X S Z D A T A U Y B
R F Q T U U T Q F R F O B
I E X E X E I T A U P O N G
P P R E S S Z W D Q D B F N P
F T E T A X O O U U E A G A
P F W F C P D R D P C N L
E I E T G S U R G E G E
T L R F D I E A L N L
S T S T T W I N G E
M Z D E D Z M O L
U K M M K U T V

The diagram illustrates the difference between a naive Viterbi decoder and a beam search decoder. It shows two versions of the same sentence: "GREETING ONHNNOROWO TDADIDODDOA COOLVLOORSWP TATAPZSTOENA OOWOOOARORSWR PCTCPOSSSSSOONC SPAOV SARARDRHIS CSCPCSONZAMNRC SPERERDRVRHAS SPCSPNONDNDC ESERERTGHAS OALNPNONRC GRIRICHIS LGPGPGOD".

On the left, the naive Viterbi decoder is shown with red lines indicating the single path of maximum local probability. This path leads to the word "OONC" in the 10th position, which is incorrect. On the right, the beam search decoder is shown with green lines indicating the top-k candidates maintained in the beam. This path leads to the word "COOD" in the 10th position, which is the correct word.

The diagram also shows the continuation of the paths for the words "OONC" and "COOD" in the next step of the decoding process, further highlighting the difference in the paths taken by the two decoders.

Результаты

Кроссворд и
описывающие поле
регулярные выражения

[PUE].P?E?(.)T\1
V..A[DU]UU[OUV]??
(P[IP])?P?EILO.?P?
.RE.[AD].D.?V.[HV]
..[LP]?G..LA.E.A?
P?Y?[GDA]RD?.[RD][RA][GD]GD?RA
EH[GJ]??.[HJL]G(..)+.?G[LJ]?.
G[GN]?..OA.DB[OD].?..?DY[PAD]
[GN]E..?GP?G(..)(.)\1\2\1\2\1
IVEB[RB](B[DB])*. [JD][RI].
TOB?.G..?(..)(.)\1\2\1
EEE.[EP]?E?[EI]B..?.
[AE]N.L....?.A
R?.T[EA]T[AT]WN?Y?
..?L[GL]?.[DG].G[LG]??

P U P P E R T R
V G M A D U U O U
P I P I E I L O L P
A R E L D R D G V Y H
A L P G U G L A Z E J A
P Y A R D G D R D G D R A
E H J G L G L G L G L G J H
G N H O A R D B D R D U D Y P
N E T C G P G L G L G L G L
I V E B R B D B D R D I J
T O B E G W T L G L G L
E E E B E E E B D I P
E N I L L N G L G A
R B T E T A W N Y
G E L D G G G L

P.?J.?PA..
[HNI]Y.?IG?I..[GY]?G
AJ.?.[DA]L.[AWJ]?L.[DW]
.[GR].U.[GRU]?...A[BG]
HJ?D(..)D\1D[LD]E?NT?D
.G?Y.GG..[BL][BT].L?.L
.?LV.?.L..D..[DLZ]T?E
R[ROA]?OG[AB].G[BG]..(..)(.)I\2\1
T[UD]..[LU].LDP.?E.?[NR]R
[RI]U.(..)(.)\2\2\1.BBE?.
E..D?U[LDE].AC?(..)O\1U?
P[AG][RI]LG[LR]?G?O(..)V\1
.[MJA]P.?P.JHE.
UGIR.Y?[LH]N*
(..)V\1AAP.G

.?RE..[IN]N.
E?B?NE.V..E?
LT[IL]E.E.?HB?..
D[EL]?..BE?..?..O.?..A?
.T.EG[ER]G(..)G\1L.
D?AN?[EP]W[BR].RL?R(..)R\1
G?.G.T(..)G\1I?[EG][DI]G.I.
L[NB]?LBL?(B..)*GU[LP]PG?..
YGD..GD.?DG.I[MI].
A[IL][IRL]?R?.R[LR].?RL?.[RE].P
...GD[DG]?D?AG?.[ID]D.
(..)I\1U?..GZ.?L[UL].
.[EG]DG.?EV[RDO].?..
[JL]..?..?..?..OT
.HAA.?PUR



Результаты

<https://github.com/GitProger/practice-regex-cross>

Мы смогли, но можно лучше

Благодарности

Мы благодарим нашего научного руководителя Дворкина Михаила Эдуардовича за научное руководство и направление теоретической части нашей работы.



Спасибо за внимание

