

Zaawansowany algorytm gry w Scrabble

Jakub Turek
J.Turek@stud.elka.pw.edu.pl

Promotor: dr inż. Jakub Koperwas

Wydział Elektroniki i Technik Informatycznych

25 kwietnia 2014

Scrabble

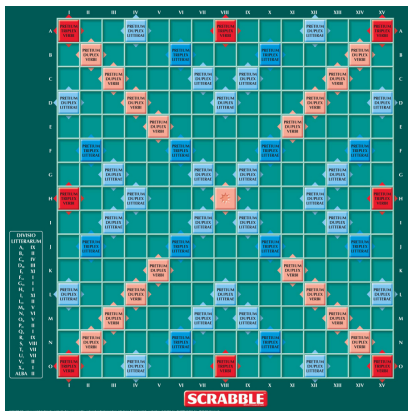
Gra słowna polegająca na układaniu na określonej planszy wyrazów z losowanych liter.

Wielki słownik ortograficzny - PWN 2003, 2006, 2008 - E. Polański

Historia Scrabble

- ▶ 1938 r. - gra *Lexiko*, Alfred Mosher Butts.
- ▶ Lata 40. - *Criss-Crossword* - udoskonalona wersja *Lexiko*.
- ▶ 1948 r. - James Burnot, *Scrabble*.
- ▶ Lata 80. - Hasbro.
- ▶ Lata 80. - teleturniej *Scrabble*.
- ▶ Obecnie - 121 krajów, 29 różnych języków.
- ▶ Do chwili obecnej - 150 milionów sprzedanych egzemplarzy.

Plansza



- ▶ Wymiary 15×15 .
- ▶ Premie:
 - ▶ literowe:
 - ▶ podwójna,
 - ▶ potrójna.
 - ▶ słowne:
 - ▶ podwójna,
 - ▶ potrójna.
- ▶ Środek planszy:
 - ▶ pierwszy wyraz musi przechodzić przez pole,
 - ▶ podwójna premia słowna.

Płytki (polska wersja)

| | | | | | | | | |
|--------|--------------|----------------|---|---|---|---|---|---|
| Litera | Ilość płytek | Liczba punktów | A | 9 | 1 | M | 3 | 2 |
| | | | Ą | 1 | 5 | N | 5 | 1 |
| | | | B | 2 | 3 | Ń | 1 | 7 |
| | | | C | 3 | 2 | O | 6 | 1 |
| | | | Ć | 1 | 6 | Ó | 1 | 5 |
| | | | D | 3 | 2 | P | 3 | 2 |
| | | | E | 7 | 1 | R | 4 | 1 |
| | | | Ę | 1 | 5 | S | 4 | 1 |
| | | | F | 1 | 5 | Ś | 1 | 5 |
| | | | G | 2 | 3 | T | 3 | 2 |
| | | | H | 2 | 3 | U | 2 | 3 |
| | | | I | 8 | 1 | W | 4 | 1 |
| | | | J | 2 | 3 | Y | 4 | 2 |
| | | | K | 3 | 2 | Z | 5 | 1 |
| | | | L | 3 | 2 | Ż | 1 | 9 |
| | | | Ł | 2 | 3 | Ź | 1 | 5 |

- ▶ 98 płytek z literami.
- ▶ Każda litera ma przyporządkowaną punktację.
- ▶ Ilość płytek proporcjonalna do częstotliwości występowania litery.
- ▶ Punktacja odwrotnie proporcjonalna do częstotliwości występowania litery.
- ▶ 2 *blanki*.

Reguły

- ▶ Na stojaku 7 wylosowanych płytek.
- ▶ Naprzemienne ruchy.
- ▶ Prawidłowy ruch:
 - ▶ Płytki wyłożone w jednym wierszu (lub kolumnie) w sposób ciągły.
 - ▶ Wykorzystanie przynajmniej jednej litery znajdującej się już na planszy.
 - ▶ Tworzy poprawny wyraz czytany od lewej do prawej (lub od góry do dołu).
 - ▶ Wszystkie płytki przylegające tworzą poprawne wyrazy w układzie krzyżówkowym.
- ▶ Koniec gry - pierwszy gracz, który nie ma płytek na stojaku.
- ▶ Wygrywa gracz z największą liczbą punktów.

Dopuszczalne słowa

Dopuszczalne jest wykorzystanie wszystkich słów znajdujących się w dowolnym słowniku języka polskiego (wraz z poprawnymi odmianami) z wyłączeniem:

- ▶ nazw własnych (wyrazów pisanych z wielkiej litery),
- ▶ skrótów,
- ▶ przedrostków, przyrostków,
- ▶ wyrazów wymagających użycia apostrofu lub łącznika.

Słownik wyrazów do gier

Słownik wyrazów do gier to lista wszystkich słów, wraz ze wszystkimi poprawnymi odmianami, dopuszczalnych do wykorzystania w grach słownych.



Porównanie OSPS i słownika alternatywnego

| | OSPS | Słownik alternatywny |
|---------------------|--|--|
| Wydawca | Polska Federacja Scrabble, Polskie Wydawnictwo Naukowe | Serwis z grami online Kurnik |
| Przeznaczenie | Gry turniejowe | Gra „Literaki” |
| Liczba słów | 2 477 212 | 2 703 830 |
| Źródło | Zamknięta lista słowników języka polskiego, ortograficznych, wyrazów obcych wydawnictwa PWN | Otwarta lista słowników języka polskiego, ortograficznych, wyrazów obcych |
| Przykładowe różnice | <ul style="list-style-type: none">▶ basfu▶ gral▶ meru▶ noblów▶ późnmyż▶ szwed | <ul style="list-style-type: none">▶ aeolipile▶ donowi▶ feroce▶ geez▶ tyiyn▶ żad |

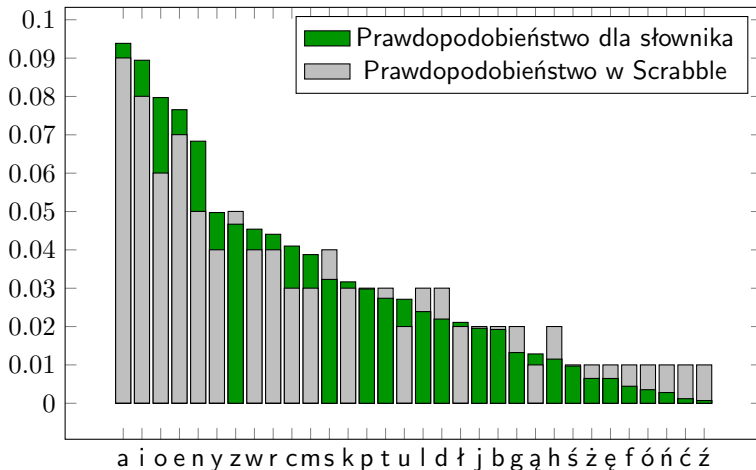
Dalsze rozważania przeprowadzane będą dla
słownika alternatywnego.

Statystyczny opis słownika

Na podstawie statystycznego opisu słownika można wyprowadzić szereg użytecznych heurystyk:

- ▶ Częstotliwość występowania liter.
- ▶ Najbardziej prawdopodobne n-gramy.
- ▶ Najlepsze otwarcia.
- ▶ Najlepsze kombinacje liter.

Prawdopodobieństwo występowania liter



Prawdopodobieństwo występowania bigramów

N-gram

Sekwencja składająca się z n liter, znaków lub wyrazów.

- ▶ Unigram.
- ▶ Bigram.
- ▶ Trigram.
- ▶ 4-gram.
- ▶ ...
- ▶ N-gram.

| Bigram | Wystąpienia |
|--------|-------------|
| ni | 1 077 436 |
| ie | 1 028 249 |
| ow | 645 018 |
| an | 507 205 |
| wa | 484 295 |
| za | 313 370 |
| po | 301 636 |
| ch | 296 749 |
| ał | 294 734 |
| ia | 284 247 |

Prawdopodobieństwo występowania n-gramów

| Trigram | Wystąpienia |
|---------|-------------|
| nie | 635 196 |
| owa | 307 277 |
| ani | 195 186 |
| wan | 180 460 |
| cie | 148 513 |
| nia | 142 201 |
| jąc | 131 792 |
| prz | 130 283 |
| wał | 126 134 |
| rze | 116 370 |

| 4-gram | Wystąpienia |
|--------|-------------|
| owan | 127 626 |
| ował | 88 130 |
| wani | 78 095 |
| niep | 77 449 |
| prze | 73 230 |
| ując | 67 062 |
| ania | 61 398 |
| ając | 59 499 |
| ście | 56 462 |
| łaby | 55 380 |


| 5-gram | Wystąpienia |
|--------|-------------|
| owani | 54 991 |
| niepo | 40 329 |
| ałaby | 37 581 |
| yście | 33 161 |
| owała | 28 175 |
| niewy | 26 193 |
| owane | 25 555 |
| wania | 25 551 |
| owany | 25 542 |
| ałyby | 25 282 |

Prawdopodobieństwo występowania n-gramów (2)

| 6-gram | Wystąpienia |
|--------|-------------|
| owania | 17 609 |
| wałaby | 17 161 |
| byście | 16 821 |
| liście | 15 910 |
| aniami | 15 585 |
| aniach | 15 585 |
| łyście | 15 405 |
| owanie | 14 674 |
| owałby | 14 437 |
| nieprz | 14 328 |

| 7-gram | Wystąpienia |
|---------|-------------|
| owałaby | 12 401 |
| libyśmy | 12 267 |
| łybyśmy | 12 094 |
| ałyście | 9 902 |
| aliście | 9 304 |
| ibyście | 8 445 |
| libyści | 8 443 |
| ybyście | 8 317 |
| łybyści | 8 314 |
| nieprze | 8 111 |

Najlepsze otwarcia

| | | | | | | | | | | | | | |
|---|---|----|---|---|---|---|---|---|---|----|---|---|-------------|
| | | x2 | | | |  | | | | x2 | | | |
| P | Ó | Ż | N | O | Ś | Ć | | | | x2 | | | 126 punktów |
| | | x2 | | | | B | Ł | Ą | D | Ż | Ż | E | 124 punkty |
| | | x2 | | | | U | B | O | D | Ż | Ż | E | 124 punkty |
| | | x2 | | | | P | Ó | J | D | Ż | K | Ę | 124 punkty |
| | | x2 | | | | G | Ł | Ó | D | Ż | Ż | E | 124 punkty |
| | | x2 | | | | U | B | Ą | D | Ż | Ż | E | 124 punkty |
| | | x2 | | | | U | G | Ó | D | Ż | Ż | E | 124 punkty |
| | | x2 | | | | B | L | U | Ż | Ń | Ż | E | 124 punkty |
| | | x2 | | | | P | Ó | J | D | Ż | K | Ą | 124 punkty |
| | | x2 | | | U | G | R | Z | Ą | Ż | Ć | | 124 punkty |

Najlepsze kombinacje liter

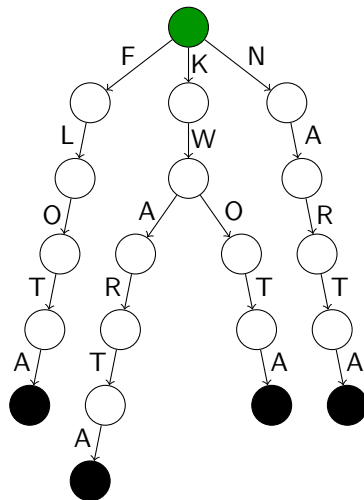
Najlepsze kombinacje liter to zawartość stojaka, która umożliwia ułożenie (niezależnie) jak największej ilości słów.

| 6 liter | Kombinacje |
|------------------|------------|
| e, m, n, o, r, t | 10 wyrazów |
| a, i, k, l, n, o | 10 wyrazów |
| a, e, i, l, m, n | 10 wyrazów |
| e, i, k, m, o, s | 9 wyrazów |
| a, i, k, m, n, o | 9 wyrazów |
| a, i, l, m, o, s | 9 wyrazów |
| a, i, k, o, t, w | 9 wyrazów |
| a, i, k, n, t, u | 9 wyrazów |
| a, e, k, l, s, z | 9 wyrazów |
| a, e, i, k, m, r | 9 wyrazów |

| 7 liter | Kombinacje |
|---------------------|------------|
| e, i, k, l, n, o, w | 12 wyrazów |
| a, e, i, k, p, r, s | 12 wyrazów |
| a, e, i, k, l, n, p | 12 wyrazów |
| a, e, k, n, r, t, y | 11 wyrazów |
| a, i, k, m, o, p, s | 11 wyrazów |
| a, i, k, m, o, r, w | 11 wyrazów |
| a, a, i, k, l, m, s | 10 wyrazów |
| a, i, k, m, o, s, t | 10 wyrazów |
| a, i, k, l, n, o, w | 10 wyrazów |
| a, i, k, l, m, n, o | 10 wyrazów |

Wyznaczanie wszystkich legalnych ruchów

- ▶ Algorytm opisany w pracy *The World's Fastest Scrabble Program* A. W. Appela i G. J. Jacobsona.
- ▶ Algorytm z nawrotami.
- ▶ Bazuje na skompresowanej, grafowej odmianie drzewa *trie* o nazwie **DAWG** (ang. **D**irected **A**cyclic **W**ord **G**raph).



Algorytm Appela-Jacobsona (1)

1. Redukcja złożoności problemu do jednego wymiaru:

- ▶ rozpatrywanie ruchów wyłącznie poziomo,
- ▶ ograniczenie zbioru wyłącznie do jednego wiersza.

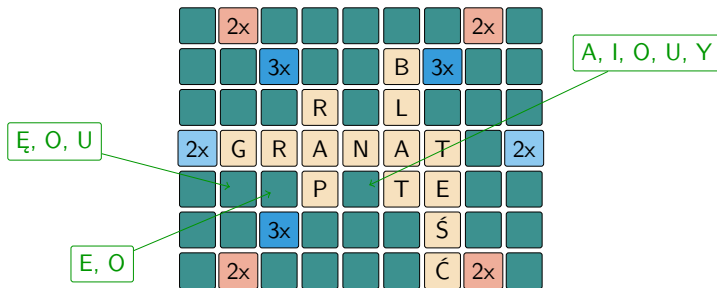
Rozumowanie należy powtórzyć dla wszystkich wierszy, a następnie transponować planszę i zastosować do ruchów w pionie.



Algorytm Appela-Jacobsona (2)

2. Ograniczenie zbioru znaków możliwych do wstawienia w miejsce pustych płytek:

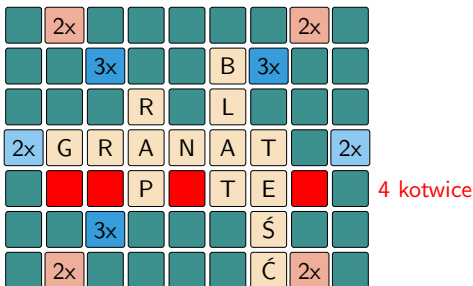
- ▶ ruch w danym kierunku może skutkować tworzeniem nowych słów w kierunku przeciwnym,
- ▶ słowa utworzone w kierunku przeciwnym powstają zawsze poprzez dodanie jednego znaku.



Algorytm Appela-Jacobsona (3)

3. Wyznaczenie kotwic (ang. *anchors*):

- ▶ kotwica to najbardziej wysunięta na lewo płytki nowego słowa, która jest przyległa do płytki istniejącego już na planszy słowa,
- ▶ kotwicą może być każde puste miejsce przyległe do płytki znajdującej się na planszy.



Algorytm Appela-Jacobsona (4)

4. Rozwinięcie słów, wychodząc od wyznaczonych kotwic, z uwzględnieniem ograniczeń.

Lewa strona

- ▶ Obejmuje wszystkie płytki na lewo od kotwicy.
- ▶ Może:
 - ▶ Składać się wyłącznie z płytek już znajdujących się na planszy - przypadek trywialny.
 - ▶ Składać się wyłącznie z płytek znajdujących się na stojaku. Wymaga wyznaczenia wszystkich możliwych kombinacji płytek.
 - ▶ Być pusta.

Prawa strona

- ▶ Obejmuje kotwicę oraz wszystkie płytki na prawo od niej.
- ▶ Wyznaczana poprzez dopełnianie lewej strony wyrazami ze słownika.
- ▶ Poszczególne litery muszą być dostępne na stojaku, a także spełniać ograniczenia wyznaczone dla poszczególnych pól planszy.

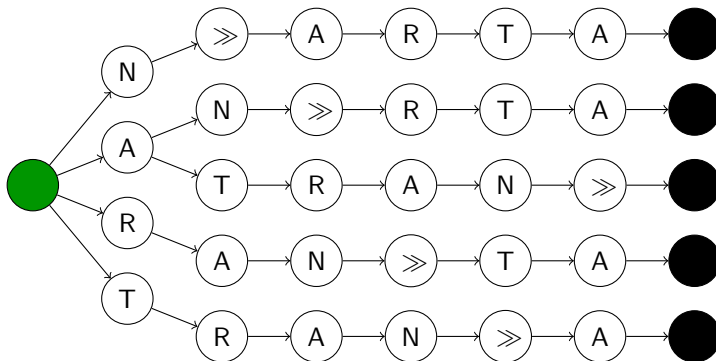
Algorytm Appela-Jacobsona - wydajność

Potencjalnym problemem wydajnościowym jest wyznaczanie wszystkich możliwych kombinacji prefiksów:

- ▶ W pesymistycznym przypadku kotwica może być skrajnie prawą płytką wyrazu.
- ▶ Dla określonych liter na stojaku może istnieć do $6! = 720$ lewostronnych kombinacji do zbadania.
- ▶ W przypadku, gdy na stojaku znajdują się dwa *blanki*, liczba kombinacji rośnie do $\frac{4! \times 32^2}{2} = 12288$.
- ▶ Nadmiarowość obliczeń - duża część badanych kombinacji może nie istnieć (lub nie posiadać rozwinięć) w słowniku.

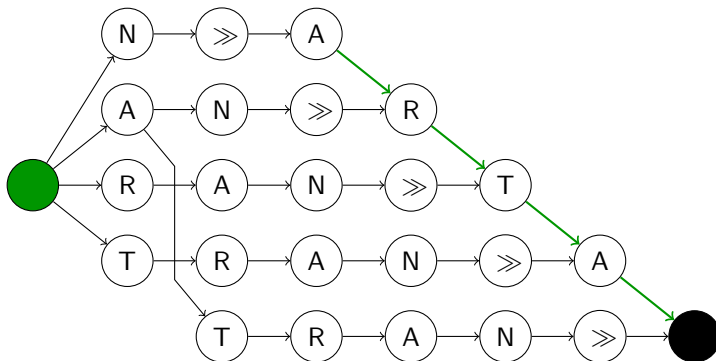
GADDAG

- ▶ S. A. Gordon, *A Faster Scrabble Move Generation Algorithm*.
- ▶ Struktura nastawiona na szybkie prefiksowanie wyrazów.



GADDAG - wady

- ▶ Duża złożoność pamięciowa.
- ▶ Można próbować minimalizować graf po węzłach zawierających \gg .



Maven i Quackle - porównanie

| | Maven | Quackle |
|--------------------------|---|---|
| Autorzy | Brian Sheppard | Jason Katz-Brown, John O'Laughlin |
| Źródło | Zamknięte | Otwarte (C++, Qt) |
| Struktura słownika | DAWG | GADDAG |
| Strategia | Zależna od fazy gry. Wykorzystanie heurystyk i symulacji do ewaluacji najbardziej korzystnych ruchów. | Zależna od fazy gry. Wykorzystanie heurystyk i symulacji do ewaluacji najbardziej korzystnych ruchów. |
| Wyniki przeciwko ludziom | ▶ 9-5 vs Adam Logan (1997) ▶ 6-3 vs Joel Sherman (2006) | ▶ 3-2 vs David Boys (2006) |
| „Bezpośrednie” starcie | ▶ 30-6 | ▶ 32-4 |

„It's still better to be a human than to be a computer” - David Boys

Fazy gry

1. **MG** - mid-game:

- ▶ Trwa od momentu rozpoczęcia gry, aż do osiągnięcia fazy pre-endgame.

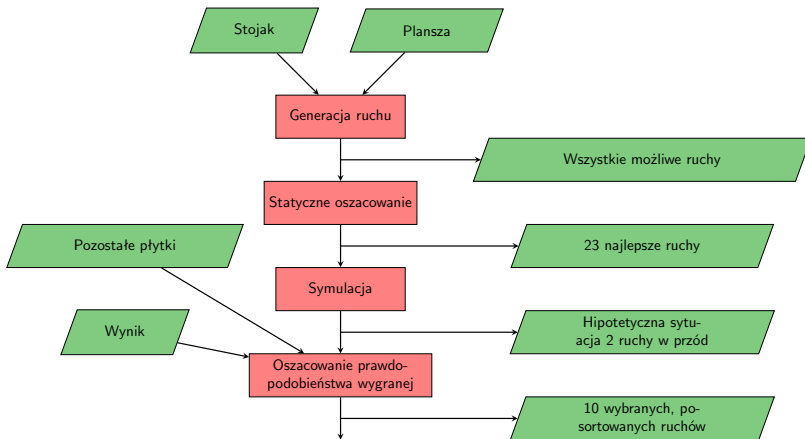
2. **PEG** - pre-endgame.

- ▶ Dzieli się na dwa etapy - PEG-1 oraz PEG-2.
- ▶ Występuje, gdy do pobrania pozostają odpowiednio jedna lub dwie płytki.
- ▶ Przez PEG przechodzi ponad połowa gier.

3. **EG** - endgame.

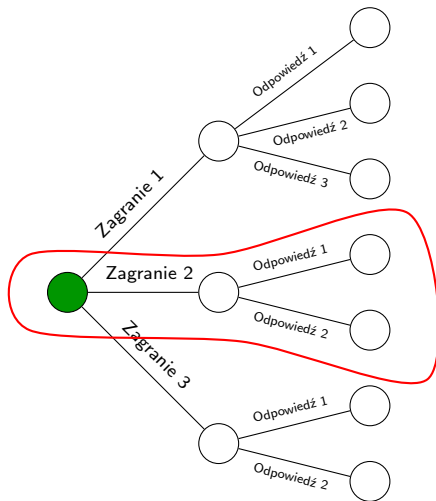
- ▶ Rozpoczyna się, gdy pobrane zostaną wszystkie płytki.
- ▶ Wiadomo jakimi literami dysponuje przeciwnik.

Mid-game (Quackle)



(Pre-)Endgame

- ▶ W fazach PEG, EG możliwe jest zastosowanie wyszukiwania wyczerpującego przestrzeni stanów.
- ▶ Algorytmy przeszukiwania $\alpha - \beta$, A^* , B^* .
- ▶ Obliczenia progresywne - przeszukiwanie rozpoczynane w miejscu, w którym można podjąć szybką i pewną decyzję.



Dziękuję za uwagę!