CROSSWORDS GENERATOR

VASILE CIOBAN, VIRGINIA NICULESCU, VASILE PREJMEREAN

Abstract. Crosswords represent a form of an intellectual game that is widespread throughout the world. If in many countries it is used only as a mild form of fun, in Romania is regarded as one great game, with many magazines dedicated to this fascinating game of mind. This paper deals with squares of words without black boxes, and the goal is to generate all possible solutions of a specified size, using a given list of words. A search algorithm is proposed, which is based on using prefixes of the words from the given words list. The experiements use lists of words in Romanian, but the proposed algorithm could also work with a list of words from any other language, or with a list of terms from various domains.
Keywords: crosswords, generator, search, complexity, performance

1. Short history

1.1. **Pompey, the first anonymous grid.** The first crossword appeared on a wall in the ancient city of Pompei. It's a magic square whose words are read in all directions: from right to left, from left to right, from top to bottom, from bottom to top. Their meaning does not change, because the order of words does not matter much in Latin. The approximate meaning of this phrase is as follows: "The Arepo worker uses the wheels as a form of work." Crosswords in the modern form, a

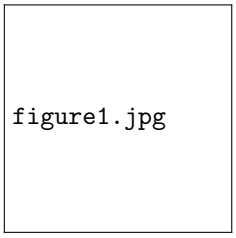

Figure 1. The first magic square drawed on a Pompei wall.

game appreciated all over the world, met in December 2013 a century of existence. On December 21, 1913, an American violinist published the first crossword grid. Since then, this game has seen a real triumph throughout the world. The reasons for the great success that crosswords had in the last century are analyzed in an article devoted to this phenomenon on the site of the magazine *Le Nouvel Observateur* [1].

1.2. **Pioneering.** With the idea of including black boxes in a square, an American-born British violinist, Arthur Wynne, can be considered as the true inventor of crosswords. He proposed his first grid in Germany without being successful. On December 21, 1913, Arthur Wynne published it in the New York World newspaper (in his supplement called Fun), but without black boxes, which he later introduced. Other publications came immediately:

- England: Morley Adams, in 1924, in Sunday Express;
- France: in the same year (on the front page of the *Dimanche-Illustre* newspaper, then in other publications, such as *Le Gaulois* and *l'Excelsior*).

Tristan Bernard, the famous French writer, has consolidated the success of the game by introducing small drawings into his grids as black boxes. When they contained too many clues, his grids were accompanied by explanations.

1.3. **Records.** We recall some recordings over time. The French daily *Le Nouvel Observateur* published in 1989 a 7×7 grid (7 rows and 7 columns) without any black boxes. *The Impossible Magazine* published an 8×8 grid without any black box created by Jean-Jacques Salgon. In 2010, Claude Coutanceau created a 9×9 grid with no black box and no reapeted word.

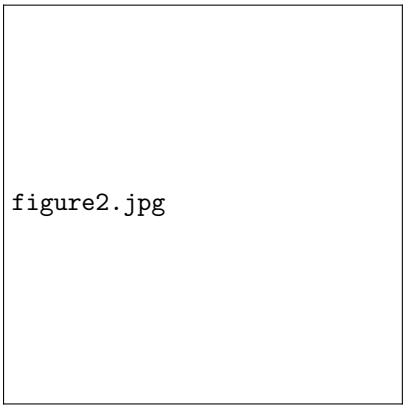

FIGURE 2. The 9×9 grid with no black box and no reapeated word, created by Claude Coutanceau.

In 1997, Didier Clerc and Pierre-Claude Singer built a grid of 160,000 squares and 50,139 words, (with black boxes) which was included in the Guinness Book of Records [3]. The largest crossword puzzle in the world was exposed on the wall of a residential building in Lviv, Ukraine, the event being part of a tourism promotion campaign in this city, included in UNESCO World Heritage in 1998 [4]. This crossword puzzle has some specific characteristics:

: - the crossword puzzle has a height of approximately 30.5 meters;
: - the giant crossword was created to attract tourists to the city, but the words are written in the Cyrillic alphabet;
: - visitors are encouraged to collect and decipher clues during weekend walks, exercising their brains and legs;
: - because the size of 19-by-34 boxes - is too large to be filled by hand, the artists responsible for this project have adopted a novel method of displaying the answers: during the night, the fluorescent letters in the boxes are on, in order to display solutions; amateurs can come every night to see if they have found the right answer;
: - clues for completing the storyline are spread around the city, around the main attractions, including parks, fountains and theaters.

In Romania, the linguist Dr. Nicolae Andrei published a square of 9x9 [5], with no black boxes, reproduced in the following figure:

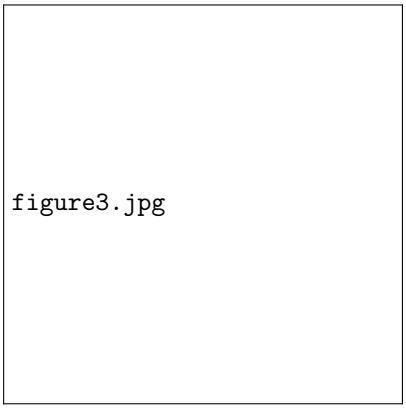

FIGURE 3. The 9x9 network with no black box and no reapeated word, created by Nicolae Andrei.

1.4. **Evolution of the black boxes.** (for the words' delimitation) While the Anglo-Saxon school has many black boxes in symmetrical positions, the French school places them where it is needed, trying to limit their number. The more difficult the grids, the fewer black boxes are needed to allow as many crosswords as possible. In the newspapers, where the grids address a hurried reader, the question of difficulty can not be put. In Romania, the number of black boxes is limited to 15% of the total number of boxes.

1.5. **Espionage.** Shortly before the 1944 Landing, a few grids in the Daily Telegraph contained some codes that were supposed to be secrets - Utah, Omaha, Neptune, Mulberry, Overlord, Juno, Gold, Sword - that created panic among Allied troops. Crossword grids began to be considered as spy boxes. Their author, Leonard Dawe, has been investigated and arrested. Leonard Dawe was a severe professor with a rigid morality, known for his football talent and often criticized for playing soccer matches while keeping his glasses on his nose. He was accused of publishing in a grid the secret code "Dieppe" (under the definition "French Port"), just days before the Allied raids in this French city (a real military disaster). The British professor explained that his grids were too old for him to know the codename given to the beaches the Allies landed and was released. In 1984, Ronald French revealed that he was Leonard Dawe's student at the time, and that he was the one who suggested that his teacher introduced those words into the grids that he had heard from the British soldiers. The "spy" was then 14 years old.

1.6. **Romania.** In Romania the number of crossword puzzle enthusiasts is huge. More than 50 titles centered on this niche are printed monthly, and their total circulation exceeds 500,000 copies. Who did not just untie a crossword puzzle? If there is someone who did not, he/she even hit them through the newspapers and knows them all. It is a form of entertainment with educational and scientifically proven results in the proper functioning of the brain. The legend says no crosswords solver has ever suffered a stroke. The people behind these books are almost unknown. Because yes, beyond the black and white squares there is a world. And it's a fascinating world – with its own history, with achievements, with failures and especially with its pride. It is the world of the people in whose mind the words are crossed and the most spiritual definitions are born. Starting with the appearance of the first chariot in our country, in the interwar period, continuing with the explosion in Communism, when the Rebus magazine demanded 800,000 copies per month. The

story of "Mr. REBUS" After December 1913, crossword puzzles also penetrated into Europe. In 1924, the carts arrived in England and France, and in 1925 they began to appear in Romania in the magazines "Mirror of the World" and "Gazeta de sâmbătă" ("Saturday Gazette"). Nobody suspected at that time that it was the beginning of a real phenomenon. In May 1926, "Saturday Gazette" invites its colleagues to a crossword contest, the first organized in Romania. The contest was successful and was followed by others, such that in 1928 a country competition was organized, after which the young student Gheorghe Popescu (nicknamed "Rebus" because of his passion for enigmistic games) was declared national champion at crosswords. He composed a 6-black boxes and 12-long-old chunk, which became a true record in the guild. In 1929, the popularity of the young Gheorghe Popescu as a promoter of crossword puzzles reached its peak, being known in the period under the name "Mr. Rebus".

1.7. **Emphasis on instructive character.** In 1931, "Hertz" publishers launched "a small gaming book" called the "Games Magazine". Indignant that others are taking advantage of the expansion of the phenomenon they have created, the *Adevărul* publishing house urges Popescu-Rebus to compose the model of a similar magazine to make the competition goes. After a few days, he presented himself to the head of the publishing house with a draft, proposing that the new publication be named *Enigma*. Aware of the young man's popularity, the head of the publishing house decided: "No! It will be called *Rebus*". It is the original moment of a "brand" around which the crossword puzzle enthusiasts in Romania over the last nine decades gravitated, but also of a semantic confusion which has survived to this day: for the crossword puzzle, the name of "rebus", which was wrong because "rebus" defines, in fact, a species of enigmistic games. In June 1931, the first issue of the 16-page *Rebus* magazine appeared. It contained crosswords, charades, riddles, chess, bridge, cinema, sports, curiosities, graphology, perspicacity, etc. "Since its first issue, Rebus has made its point of view on gaming, firmly emphasizing its instructive character, in contrast to the similar headlines and magazines in the country and abroad that saw in these forms of entertainment a mere passage of time" wrote Popescu-Rebus later in *The History of Romanian Rebusism*. In the wake of this belief, the magazine had the idea of concluding a partnership with the Ministry of Education, on the basis of which it organizes inter-school and public contests in every number of books in which the material from the textbooks was introduced according to the curricula. "Through this innovation, Rebus magazine definitively set itself on an original position, as an auxiliary of modern education, as a mentor and friend of the youth," wrote her founder. The publication aimed at giving the school-age reader a "general culture, to develop his vocabulary, to train him with independent work skills", the teaching books being assumed as a hundred percent Romanian innovation. The reluctance of the Romanian rebellion movement has become a stumbling block in the increasingly dramatic succession of the historical events of those times. On January 1, 1939, the appearance of the magazine *Rebus* was stopped and with it the rebellion from us in the country was going to enter into a long sleep. In this context, Gheorghe Popescu-Rebus withdrew to his native village Pietrari (Vâlcea), where he worked as a teacher and lived until the beginning of 2000, reaching the venerable age of 92 years. After the World War II Romanian rebusism was again one of the first in the world.

The real moment of rebirth was the summer of 1957, when the *Rebus* magazine was re-established, under the leadership of Ion Pătraşcu. As a symbolic gesture, Pătraşcu asked for the blessing of Gheorghe Popescu-Rebus, retired to his native Vâlcea, and promised to continue the work he had begun in the interwar period. Indeed, the magazine *Rebus* has assumed an educational role, raised the level of difficulty of the carts and created a huge community around it. It had to make compromises, the Party demanded communist-themed books in every number (steel, agrarian, etc.),

and censorship checked to avoid hiding hidden messages against the regime, but people passed over them and turned it into - one of the most popular publications of those times. The request was 800,000 copies at one time, but they did not allow us to exceed 100,000, saying that we just do not want to print more than *Scanteia*, says Alexandru Păsărin (over 80 years old at the magazine *Rebus* for over 55 years!), who, before being the director of the magazine, is the man who advertised this publication. Alexandru Păsărin walked and promoted the "rebus" in school, organized school reunion camps, moved to the province and laid the foundations of circus circles that were set up beside various institutions (cultural houses, houses of the army, etc.). These circles were organized, periodically, bi- or tri- lateral matches, symposiums or simple agape rebus. He said: "25 years I walked through the country to set up these circles in dozens of cities".

1.8. **Innovation of "spiritual" definitions.** (definition crosswords puzzle) From this formidable competition for winning a place in the pages of the magazine were born more and more valuable books. In order to overcome their rivals, the rebels had to innovate. The so-called "metaphorical definitions", a Romanian innovation, appeared, says Alexandru Păsărin: "The foreigners do not have this: they only go to synonymy, that is, they give the definition of love and the answer is love. In our case, the first definition that shocked was a Bucharest rebusist, who gave the definition of "ironing iron". Everyone was thinking of ironing, but the loosening was "blacheu" (a piece of iron that is applied to the tip or shoe of the shoe). It was very much appreciated and all the people started to look for definitions of this kind. People did not have free time alternatives like today, and then they were standing and thinking for hours on how to define a word as much original as possible. Thus, according to the guild, Romanian rebusism has reached the top spot in the world.

2. Some definitions

We'll give some definitions to fix the ideas.

- **Simple square** = crosswords square with no black boxes and the words are read in two directions: from left to right and from up to down.
- **Symmetrical square** = square with identical horizontal and vertical (symmetrical)
- **Complet square** = square with words in 4 directions ($\leftrightarrow$?i up $\leftrightarrow$ down) **Example:** the square from ancient city Pompei, see the Figure 1 (Section 1).

3. Problem specification

In order to find the possible solutions of a crosswords grid, we propose the development of a software application. The specification of such a software application is the following: *Input*: $n \in N^*$, L – a DEX word list, each word having the length n. (DEX = the explanatory dictionary of the Romanian language). *Output*: the list of all possible n-squares (simple / symmetrical / complete) with words from L.

A direct implementation of this program could be thought based on a "brute force" method.

- The "brute force" method implies passing all the words (from L) horizontally and checking vertically (on columns). In this case the first vertically word is taken from L, sequentially. So, if the size of the given list L is not very small, the searching solution space become enormous.

Because of this very big time-complexity, we proposed an alternative:

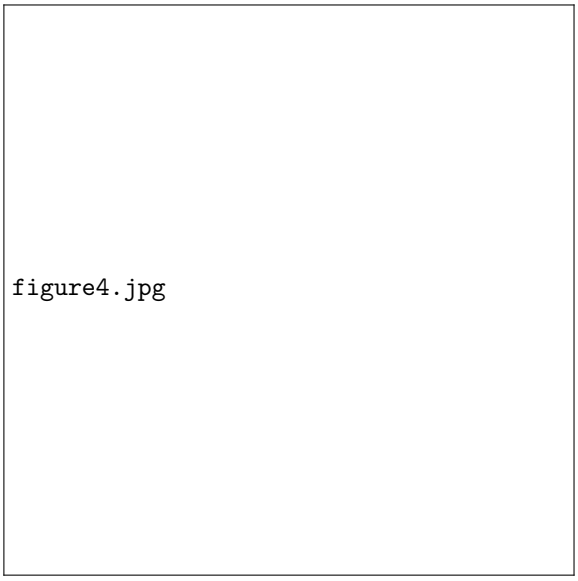

FIGURE 4. Simple squares for $n = 5$ and $n = 6$ (n is the words length).

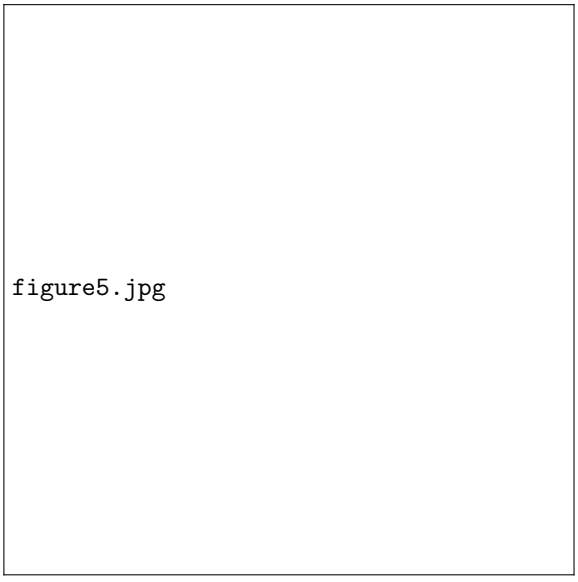

FIGURE 5. Symmetrical squares for $n = 5$ and $n = 6$ (n is the words length).

- Alternative: start from a given vertical word in the first column, then scroll through all the words that have the prefix equal to the first letter of the given word and verifying it vertically.

The program solution is based on the following: (Let n be the size of the grid - quadratic matrix of order n.)

- Search with prefixes:
  - Let be $C = (c_0, c_1, ...c_{n-1})$, the word positioned on the first column of the grid; letters $c_0, c_1, ...c_{n-1}$ are prefixes for vertical words, for each row;
  - The solutions are formed from $X_0, X_1, ..., X_{n-1}$ which are words from L with the following conditions:

    $$X_i \in D_i$$

    where

    $$D_i = \{ \text{ the set of all the words that start with } c_i\}$$

    * $x_{0,j}x_{1,j}x_{2,j}...x_{n-1,j}$ (are concatenated letters) form a word from L on any column $j = 1..n-1$ (the first column - the column 0 determines the rest);
    * the continuation conditions are given by checking the admissible prefixes (can be computed in parallel for each column; prefixes with $2, 3, 4$ or 5 letters)
- Analyze
  - for $n \leq 10$ the depth of the corresponding search tree is relatively small;
  - for computing D_i the time-complexity is great
    * for the test of words starting with the letters given in the first column the number of verifications is

      $$|D_0| * |D_1| * ... * |D_{n-1}|$$

      (for example for $n = 10$ and $|D_i| = 1000$ the number of variants are equal to $1000^{10} = 10^{30}$)
  - the "bad" variants are removed as quickly as possible - from the first levels (eg. level 2 when we have 2-character prefixes, or 3 ...)
  - continuation conditions can be verified by acceptable (possible) prefixes or negative prefixes that are not among the acceptable:
    * Which are fewer: acceptable or unacceptable prefixes?
    * Total prefixes are 26k, for example for k = 2 all prefixes are 26 * 26 = = 676 (26, alphabet letters);
    * Conclusion: the acceptable prefixes are fewer (they are taken from L).

4. Implementation and Practical Achievements

We have implemented the presented searching algorithm based on prefixes. The used language is C++ [6]. The program is parameterized by the length of words (n). For now, we have been only used the prefixes of length $k = 2, 3, 4, 5$ with the retention of their lists in the internal memory. The DEX words collection required also supplementary manipulation (arrangements). **Results:**

- for $n = 5$, we generated over 400,000 simple and symmetric squares (for an incomplete L list of about 3000 words of length 5 from DEX).
- for $n = 6$, we found over 10.000 simple and symmetric squares (for an incomplete L list of about 3000 words of length 6 from DEX)
- for $n = 8$, we found a few hundreds squares for an incomplete L list of about 25.000 DEX words of length 8.
- for $n = 9$ we found the mentioned item in page 2;
- for $n = 10$ we did not find any square on an incomplete L list of DEX words of length 10.

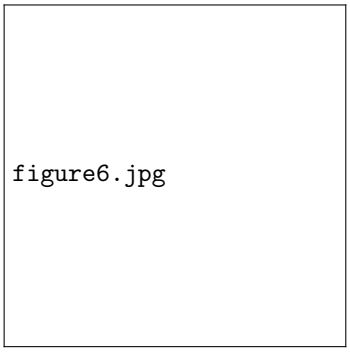

FIGURE 6. An example of an obtained $6x6$ square.

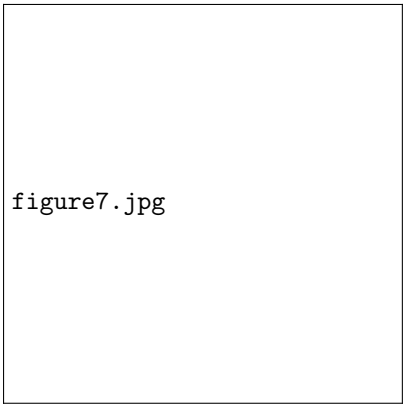

FIGURE 7. An example of an obtained $8x8$ square.

5. Conclusion and Further Work

This paper presents a preliminary work related to crosswords automatic generation. The future work goal is to better structure the possible solutions such that to obtain a general program, which could provide, in a reasonable time, solutions for squares of different sizes using a given dictionary of terms. The considered dictionaries could address different kind of information (not just literary words), and so the results could also be used in other domains. Since the estimated time-complexity of such a general program using just sequential computation is very big, we will investigate possible parallelizations of the computation. Another direction of the further work is related to the automatic generation of crossword squares that have black boxes, too.

References

[1] *Le nouvelle Observateur*. www.nouvelobs.com.
[2] *The Telegraph*. www.telegraph.co.uk.
[3] *Guinness Book of Records*. http://www.guinnessworldrecords.com
[4] *UNESCO World Heritage Centre*. https://whc.unesco.org
[5] Nicolae Andrei, Indreptar rebusistic, Editura Sport Turism, 1980.

[6] B. Stroustrup. *Programming – Principles and Practice Using C++ (Second Edition)*. 2014, Addison-Wesley, 1312 pages.

Department of Computer Science, Babeş-Bolyai University, Cluj-Napoca
Email address: vcioban@cs.ubbcluj.ro