Univerzita Komenského v Bratislave

Fakulta Matematiky, Fyziky a Informatiky

Mgr. Matúš Palmovský

Autoreferát dizertačnej práce

Descriptional complexity of operations on prefix-free languages

na získanie akademického titulu doktor

(philosophiae doctor)

v odbore doktorandského štúdia : 9.1.9 Aplikovaná matematika

Košice, júl 2017

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia na Matematickom ústave Slovenskej akadémie vied v Košiciach.

Doktorand: Mgr. Matúš Palmovský
Matematický Ústav
Slovenskej akadémie vied
Grešáková 6, 040 01 Košice

Školiteľ: RNDr. Galina Jirásková, CSc.
Matematický Ústav
Slovenskej akadémie vied
Grešáková 6, 040 01 Košice

Oponenti: Prof. RNDr. Viliam Geffert, DrSc.
Ústav informatiky
Prírodovedecká fakulta
Univerzita P. J. Šafárika v Košiciach
Košice

prof. Dr. rer. nat. Markus Holzer
Institut für Informatik
Justus-Liebig-Universität Giessen
Arndstraße 2, D-35392 Giessen
Germany

doc. RNDr. Gabriela Andrejková, CSc.
Ústav informatiky
Prírodovedecká fakulta
Univerzita P. J. Šafárika v Košiciach
Košice

Obhajoba dizertačnej práce sa koná o 23.8.2017 o 12:45 na Matematickom Ústave Slovenskej akadémie vied v Bratislave
pred komisiou pre obhajobu dizertačnej práce v doktorandskom Študijnom programe Aplikovaná matematika.

Predseda spoločnej odborovej komisie:
Prof. RNDr. Anatolij Dvurečenskij, DrSc.
Matematický Ústav
Slovenskej akadémie vied
Štefániková 49, 814 73 Bratislava

1 Introduction

Regular languages and finite automata are among the oldest topics in computer science and formal languages theory. They were examined already in 1950s and 1960s. Despite their simplicity, some problems remain open. The most challenging is the question of how many states are sufficient and necessary in the worst case for two-way deterministic automata to simulate two-way nondeterministic automata. The problem is connected the well known open problem whether or not DLOGSPACE = NLOGSPACE which is an analogous problem to P vs. NP.

Finite automata have applications in software engineering, programming languages, and other areas in computer science. But they are also important from the theoretical point of view [9]. Recently, various topics concerning this class have been investigated. One of such topics is descriptional complexity. It studies the cost of description of languages by formal systems.

Rabin and Scott in 1959 [28] described the "subset construction" and showed that every n-state nondeterministic automaton (NFA) can be simulated by a deterministic finite automaton (DFA) of at most 2^n states. Lupanov [22] then proved that this upper bound is tight by describing languages accepted by an n-state NFA whose equivalent DFA requires 2^n states. In the western literature, witness languages meeting the bound 2^n were described in 1971 in papers by Moore [26] and Meyer and Fischer [24].

Maslov [23] investigated the state complexity of union, concatenation, and Kleene closure. He also considered cyclic shift which is one of the hardest operation; the cyclic shift of a language represented by a 5-state deterministic automaton may require up to 56 million states [20]. Yu et al. [30] were the first who studied systematically the state complexity of operations on regular languages. Their paper was followed by a paper studying state complexity of operations on finite languages [**?**]. Some special operations have also been studied: Domaratzki examined proportional removals in [3], shuffle was investigated in [2], and cyclic shift in [20].

Recently, the subclasses of regular languages have been investigated, such as, for example, prefix-free and suffix-free languages [6,8], ideal languages [25], closed languages [1], union-free languages [18]. In some subclasses, the operations may have smaller complexity, while in some others, for example, in the class of union-free languages, the complexity of operations is the same as in the general case of regular languages.

The paper [29] began the study of the complexity of combined operations. In some cases, the resulting complexity may be equal to the composition of particular complexities. However, most often, it is much smaller. Combined operations which contain complementation seem to be the most interesting [11,21].

The study of so called "magic" numbers is a stream of research where not only worst-case complexities are important, but also all values that can be obtained as a result of a complexity problem are considered. The problem was stated by Iwama et al. [10]. The authors asked what values can be obtained as the size of the minimal deterministic automaton equivalent to a given n-state nondeterministic automaton. The values that cannot be obtained in such a way are called "magic" numbers in [27]. The following research showed that there are no magic numbers in the ternary case [16], while a lot of them exist in the unary case [4]. Similar results for other operations were published in [15, 17].

We study the state complexity the Kleene closure operation on regular, prefix-free, and prefix-closed languages in the first part of this thesis. We describe a binary language accepted by an n-state DFA with k-final states meeting the upper bound $m2^n - k2^{n-1}$ for its Kleene closure. This result can be used to get a language that is hard for the Kleene closure operation on alternating finite automata [19]. In the case of prefix-free languages, there are only tree possible complexities $n, n-1, n-2$. In the prefix-closed cases, we improve a know result from the literature by decreasing the size of an alphabet used for describing the witness languages.

2 Aims

1. Summarize known results on the state complexity of operations on regular and prefix-free languages
2. Study the operation the Kleene closure on regular, prefix-free, prefix-closed languages
3. Investigate the complexity of combined operations on prefix-free languages
4. Find the exact complexity of the star-complement-star in the class of the prefix-free languages
5. Examine Kuratowski algebras generated by prefix-free languages; focus on the state complexity of the generated languages

2.1 Main Results - Kleene Closure on Regular

First, we proved that the n-state automata presented by Maslov in his 1970 paper meet the upper bound $(3/4) \cdot 2^n$ on the state complexity of Kleene closure. We provided the n-state binary automata with k final states that meet the upper bound $2^{n-1} + 2^{n-1-k}$ on the state complexity of Kleene closure. Our experimental results showed that there are no holes in the hierarchy up to $n = 9$, even in the binary case. Then we examined the Kleene closure operation on prefix-free languages. We showed that the state complexity of the Kleene closure of a prefix-free language with state complexity n may attain just three values $n-2, n-1$, and n. Finally we proved that the state complexity of Kleene closure on prefix-closed languages is $2^{n-2} + 1$, and that there is exactly one binary prefix-closed language of state complexity n such that the state complexity of its Kleene closure is 2 whenever $n \geq 3$.

2.2 Main Results - Combined operation on Prefix-free

We investigated the state complexity of combined operations with concatenation in the class of prefix-free languages. We can usually obtain a much lower state complexity for combined operations compared with the compositions of the state complexities of individual operations. However, for some cases, the state complexity of combined operations and the composition of state complexities are the same. We have examined union, intersection, and star combined with concatenation. For each of these combined operations, we obtained a tight upper bound. To prove tightness we used binary or ternary alphabets, except for $L_1^* \cdot L_2$ where we used a growing alphabet of size $n+3$. We conjecture that a binary alphabet is always optimal.

operation	prefix-free languages	$\|\Sigma\|$
$L_1 \cdot (L_2 \cup L_3)$	$m + np - 4$	3
$(L_1 \cup L_2) \cdot L_3$	$(m-2)(n-2) + (m+n-4)p + (p^2 - p + 2)/2$	2
$(L_1 \cap L_2) \cdot L_3$	$(m-2)(n-2) + p$	2
$L_1 \cdot (L_2 \cap L_3)$	$m + np - 2(n+p) + 4$	2
$L_1 \cdot L_2^*$	$m + n - 2$	2
$L_1^* \cdot L_2$	$(m-1)(2^{n-1} - 1) + 1$	$n+3$

Table 1: State complexity of combined operations on prefix-free languages; $m, n \geq 4$.

2.3 Main Results - Star Complement Star on Prefix-free

We investigated the star-complement-star operation on prefix-free languages. We proved that if a prefix-free language L is accepted by an n-state deterministic finite automaton, then the language L^{*c*} is accepted by a deterministic finite automaton of at most $2^{n-3} + 2$ states. We also proved that this upper bound is tight for every alphabet containing at least two symbols. We showed that the state complexity of the star-complement-star of a unary prefix-free language is 1, except for the language $\{a\}$, where it is 2. Our computations showed that if $n \in \{4, 5, 6, 7, 8\}$, then there exist exactly one, up to renaming input symbols, binary prefix-free language of state complexity n such that the state complexity of its star-complement-star is 2. We proved that this true for all $n \geq 4$.

$n \backslash \mathrm{sc}(L^{*c*})$	1	2	3	4	5	6	7	8	9	10	average
3	-	2	1	-	-	-	-	-	-	-	2,333
4	18	1	7	2	-	-	-	-	-	-	1.75
5	374	1	83	37	24	2	-	-	-	-	1.737
6	10374	1	1638	353	482	359	172	42	26	6	1,71
7	356623	1	47123	5259	7501	8194	8044	4450	2663	1867	1,738

Table 2: The frequencies of the complexities and the average complexity of star-complement-star on prefix-free languages in the binary case; $n = 3, 4, 5, 6, 7$.

$n \backslash \mathrm{sc}(L^{*c*})$	11	12	13	14	15	16	17	18	19	20
7	896	447	608	174	-	-	164	26	-	-

Table 3: The frequencies of the complexities 11-18 of star-complement-star on prefix-free languages in the binary case; $n = 7$.

2.4 Kuratowski algebras on prefix-free

We studied Kuratowski algebras generated by prefix-free languages under operations of positive closure and complement. For each 9 possible algebras we proved whether or not it can be generated by a prefix-free language. In each case when the Kuratowski algebra can be generated by a prefix-free language, we found a regular prefix-free generator which maximizes the complexities of all generated languages. Then we considered Kuratowski algebras generated by factor-free and subword-free languages under operations of Kleene closure and complement. In this case there are 12 possible algebras. We proved that seven of them cannot be generated any factor-free language. In all the remaining cases, we found a binary regular subword-free generator which maximizes the complexities of all generated languages.

Case	$B(L)$	State complexities	Prefix-free generator
(1)	L	1 (2)	$\emptyset$ (ε)
(2)	L, L^+	n, n	Fig. 1
(5)	$L, L^+, L^\oplus$	$n, n, 1$	Fig. 2
(6)	$L, L^+, L^\oplus, L^{\oplus+}$	n, n, n, n	Fig. 3
(8)	$L, L^+, L^\oplus, L^{\oplus+}$	$n, n, 2^{n-3}+2, 2^{n-3}+2$	Fig. **??**
(9)	$L, L^+, L^\oplus, L^{\oplus+}, L^{+\oplus}$	$n, n, 2^{n-3}+2, 2^{n-3}+2, 2^{n-3}+2$	Fig. **??**

Table 4: Prefix-free generators of Kuratowski under positive closure and complement maximizing complexities of generated languages. Cases (3), (4), and (7) can not be generated by any prefix-free language; $n \geq 5$.

Case	$E(L)$	State complexities	Subword-Free Generator
(1a)	$L, L \setminus \{\varepsilon\}$	$2, 1$	ε
(1b)	$L, L \cup \{\varepsilon\}$	$1, 2$	$\emptyset$
(2a)	L, L^*, L^+	$3, 2, 3$	$\{a\}$ *over* $\{a, b\}$
(5)	$L, L^*, L^\oplus \cup \{\varepsilon\}$	$n, n-1, 2, 1$	$\{a^{n-2}\}$ over $\{a, b\}$
(8)	$L, L^*, L^\oplus, L^{\oplus *}, L^{\oplus +}$	$n, n-1, 3, 2, 3$	$\{a, b^{n-2}\}$

Table 5: Binary subword-free generators of Kuratowski algebras under Kleene closure and complement maximizing complexities of generated languages. Cases (2b), (3a), (3b), (4), (6), (7), and (9) can not be generated by any factor- or subword-free language.

and here we have prefix-free generators

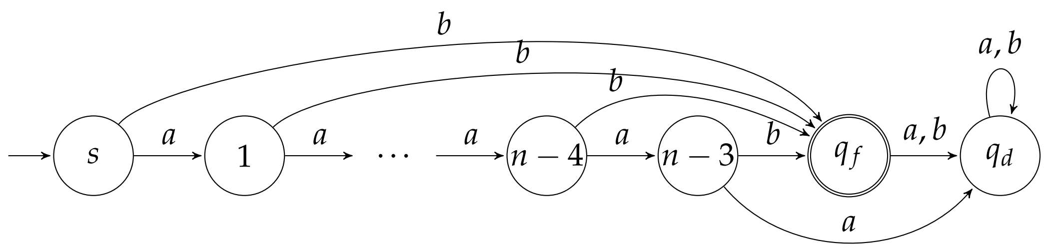

Figure 1: The binary witness DFA A for case (2) with $\text{sc}(L^+) = n$.

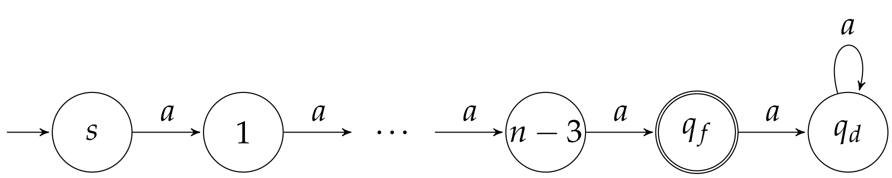

Figure 2: The unary witness DFA A for case (5).

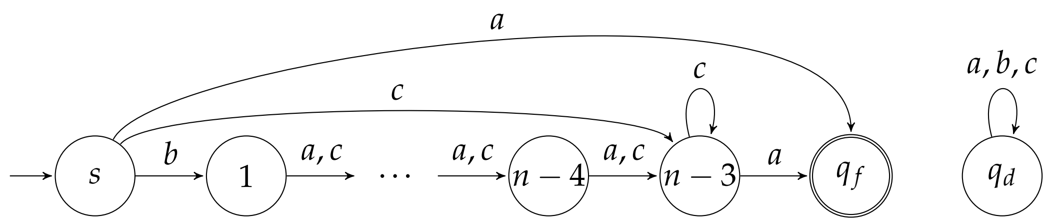

Figure 3: The DFA A for case (6); all undefined transitions go to the dead state q_d.

3 Related Published Papers (AEC):

[1] Galina Jirásková, Matúš Palmovský: Kleene closure and state complexity. In: Vinař T. (ed.) Proceedings of the Conference on Information Technologies - Applications and Theory, Slovakia, September 11-15, 2013, ITAT 2013, CEUR Workshop Proceedings, vol. 1003, pp. 94–100. CEUR-WS.org (2013)

[2] Galina Jirásková, Matúš Palmovský, Juraj Šebej: Kleene closure on regular and prefix-free languages. In: Holzer M., Kutrib M. (eds.) Implementation and Application of Automata - 19th International Conference, CIAA 2014, Giessen, Germany, July 30 - August 2, 2014. Proceedings, Lecture Notes in Computer Science, vol. 8587, pp. 226–237. Springer (2014)

[3] Galina Jirásková, Matúš Palmovský, Juraj Šebej, Peter Mlynárčik, Kristína Čevorová: Operations on automata with all states final. In: Ésik Z., Fülöp Z. (eds.) Proceedings, 14th International Conference on Automata and Formal Languages, AFL 2014, Szeged, Hungary, May 27-29, 2014, EPTCS, vol. 151, pp. 201–215.

[4] Matúš Palmovský, Juraj Šebej: Star-complement-star on prefix-free languages. In: Shallit J., Okhotin A. (eds.) Descriptional Complexity of Formal Systems - 17th International Workshop, DCFS 2015, Waterloo, ON, Canada, June 25-27, 2015. Proceedings, Lecture Notes in Computer Science, vol. 9118, pp. 231–242. Springer (2015)

[5] Hae-Sung Eom, Matúš Palmovský: Combined operations on prefix-free and suffix-free languages. In: Bordihn H., Freund R., Nagy B., Vaszil Gy. (eds.) Eighth Workshop on Non-Classical Models of Automata and Applications, NCMA 2016, Debrecen, Hungary, August 29-30, 2016. Proceedings, books@ocg.at , vol. 321, pp. 147–162. Österreichische Computer Gesellschaft (2016)

[6] Matúš Palmovský: Kleene closure and state complexity. RAIRO - Theoretical Informatics and Application 50(3), 251–261 (2016), http://dx.doi.org/10.1051/ita/2016024

[7] Jozef Jirásek, Jr., Matúš Palmovský, Juraj Šebej: Kuratowski algebras generated by suffix-, factor-, and subword-free languages. DCFS 2017.

4 Cited in

[1] Jozef Jirásek, Jr., Juraj Šebej: Kuratowski Algebras Generated by Prefix-Free Languages. In: Han, YS; Salomaa, K (eds.) Conference: 21st International Conference on Implementation and Application of Automata (CIAA) Location: Seoul, SOUTH KOREA Date: JUL 19-22, 2016. Proceedings, Lecture Notes in Computer Science , vol. 9705, pp. 150-162. (2016)

[2] Hospodár, M.; Jirásková, G.; Mlynarčík, P: Nondeterministic Complexity of Operations on Closed and Ideal Languages. In: Han, YS; Salomaa, K (eds.) Conference: 21st International Conference on Implementation and Application of Automata (CIAA) Location: Seoul, SOUTH KOREA Date: JUL 19-22, 2016. Proceedings, Lecture Notes in Computer Science , vol. 9705, pp. 125-137. (2016)

[3] Eom, Hae-Sung; Han, Yo-Sub; Jiraskova, Galina: State Complexity of Basic Operations on Non-Returning Regular Languages. FUNDAMENTA INFORMATICAE Volume: 144 Issue: 2 Pages: 161-182 Published: 2016

[4] Jozef Jirásek, Jr., Juraj Šebej: Non-regular Maximal Prefix-Free Subsets of Regular Languages. In: Brlek, S; Reutenauer, C (eds.) Conference: 20th International Conference on Developments in Language Theory (DLT) Location: Montreal, CANADA Date: JUL 25-28, 2016 Proceedings, Lecture Notes in Computer Science , vol. 9840, pp. 229-242. (2016)

5 Talks given by the author

[1] Kleene closure and state complexity. The Conference on Information Technologies - Applications and Theory, Slovakia, September 11-15, 2013, ITAT 2013

[2] Kleene closure on regular and prefix-free, prefix-closed languages. International Workshop on Černý's Conjecture and Optimization Problems, Opava, Czech Republic, November 6-7, 2014

[3] Combined operations on prefix-free and suffix-free languages. Eighth Workshop on Non-Classical Models of Automata and Applications, NCMA 2016, Debrecen, Hungary, August 29-30, 2016

References

[1] Brzozowski, J.A., Jirásková, G., Zou, C.: Quotient complexity of closed languages. Theory Comput. Syst. 54, 277–292 (2014)

[2] Câmpeanu, C., Salomaa, K., Yu, S.: Tight lower bound for the state complexity of shuffle of regular languages. J. Autom. Lang. Comb. 7, 303–310 (2002)

[3] Domaratzki, M.: State complexity and proportional removals. J. Autom. Lang. Comb. 7, 455–468 (2002)

[4] Geffert, V.: Magic numbers in the state hierarchy of finite automata. Inform. and Comput. 205, 1652–1670 (2007)

[5] Han, Y.-S., Salomaa, K.: State complexity of basic operations on suffix-free regular languages. Theoret. Comput. Sci. 410, 2537–2548 (2009)

[6] Han, Y.-S., Salomaa, K., Wood, D.: Operational state complexity of prefix-free regular languages. Automata, Formal Languages, and Related Topics, pp. 99–115. Institute of Informatics, University of Szeged (2009)

[7] Han, Y.-S., Salomaa, K., Wood, D.: Nondeterministic state complexity of basic operations for prefix-free regular languages. Fund. Inform. 90, 93–106 (2009)

[8] Han, Y.-S., Salomaa, K., Wood, D.: Operational state complexity of prefix-free regular languages. Ésik, Z.(ed.). Automata, formal languages, and related topics, pp. 99-115. Institute of Informatics, University of Szeged (2009)

[9] Hromkovič, J.: Descriptional complexity of finite automata: Concepts and open problems. J. Autom. Lang. Comb. 7, 519–531 (2002)

[10] Iwama, K., Kambayashi, Y., Takaki, K.: Tight bounds on the number of states of DFAs that are equivalent to n-state NFAs. Theoret. Comput. Sci. 237, 485–494 (2000).

[11] Jirásek, J., Jirásková, G.: On the boundary of regular languages. In: Konstantinidis, S. (ed.) CIAA 2013. LNCS, vol. 7982, pp. 208-219. Springer Heidelberg (2013)

[12] Jirásek, J., Jirásková, G.: Cyclic Shift on Prefix-Free Languages. In: Andrei A. Bulatov and Arseny M. Shur(ed.) CSR 2013. LNCS, vol. 7913, pp. 246–257. Springer Heidelberg (2013)

[13] Jirásková, G.: The Ranges of State Complexities for Complement, Star, and Reversal of Regular Languages. vol. 25, pp. 246–257. Int. J. Found. Comput. Sci. (2014)

[14] Jirásková, G.: Note on minimal automata and uniform communication protocols. In: Martin-Vide, C., Mitrana,V. (eds.) Grammars and Automata for String Processing: From Mathematics and Computer Science to Biology, and Back, pp. 163–170. Taylor and Francis, London (2003)

[15] Jirásková, G.: On the state complexity of complements, stars, and reversals of regular languages. In: Ito M., Toyama M. (eds.) DLT 2008. LNCS, vol. 5257, pp. 431–442. Springer Heidelberg (2008)

[16] Jirásková, G.: Magic numbers and ternary alphabet. Internat. J. Found. Comput. Sci. 22, 331–344 (2011)

[17] Jirásek, J., Jirásková, G., Szabari, A.: State complexity of concatenation and complementation. Internat. J. Found. Comput. Sci. 16, 511–529 (2005)

[18] Jirásková, G., Masopust, T.: Complexity in union-free regular languages. Internat. J. Found. Comput. Sci. 22, 1639–1653 (2011)

[19] Jirásková, G.: Descriptional complexity of operations on alternating and boolean automata. In Hirsch et al. [9], pp. 196-204. Available at `http://dx.doi.org/10.1007/978-3-642-30642-6-19` (2012)

[20] Jirásková, G., Okhotin, A.: State complexity of cyclic shift. , RAIRO Theor. Inform. Appl. 42, 335–360 (2008)

[21] Jirásková, G., Shallit, J.: The state complexity of star-complement-star. In: Yen, H.-C., Ibarra, O.H. (eds.) DLT 2012. LNCS, vol. 7410, pp. 380–391. Springer Heidelberg (2012)

[22] Lupanov, U.I.: A comparison of two types of finite automata. Problemy Kibernetiki 9, 321-326 (1963) (in Russian)

[23] Maslov, A.N.: Estimates of the number of states of finite automata. Soviet Math. Dokl. 11, 1373–1375 (1970)

[24] Meyer, A.R., Fischer, M.J.: Economy of description by automata, grammars and formal systems. In: Proc. 12th Annual Symposium on Switching and Automata Theory, pp. 188–191 (1971)

[25] Brzozowski, J.A., Jirásková, G., Li, B.: Quotient complexity of ideal languages. Theoret. Comput. Sci. 470, 36–52 (2013)

[26] Moore, F.R.: On the bounds for state-set size in the proofs of equivalence between deterministic, nondeterministic, and two-way finite automata. IEEE Trans. Comput. 20, 1211–1214 (1971)

[27] Iwama, K., Matsuura, A., Paterson, M.: A family of NFAs which need $2^n - \alpha$ deterministic states. Theoret. Comput. Sci. 301, 451–462 (2003)

[28] Rabin, M., Scott, D.: Finite automata and their decision problems. IBM Res. Develop. 3, 114-129 (1959)

[29] Salomaa, A., Salomaa, K., Yu, S.: State complexity of combined operations. Theoret. Comput. Sci. 383, 140-152 (2007)

[30] Yu, S., Zhuang, Q., Salomaa, K.: The state complexity of some basic operations on regular languages. Theoret. Comput. Sci. 125, 315-328 (1994)

various measures of descriptional complexity of automata, grammars, languages and of related systems. trade-offs between descriptional complexity and mode of operation.Â In a survey on descriptional complexity, Holzer & Kutrib (2010) state that "since more than a decade the Workshop on 'Descriptional Complexity of Formal Systems' (DCFS), [...] has contributed substantially to the development of [its] field of research." In a talk on the occasion of the 10th anniversary of the workshop, Dassow (2009) gave an overview about trends and directions in research papers presented at DCFS. History of the workshop. Since 2006, the Chair of the Steering Committee of the DCFS workshop series is Giovanni Pighizzini. International Workshop on Descriptional Complexity of Formal Systems. DCFS 2015: Descriptional Complexity of Formal Systems pp 231-242 | Cite as. Star-Complement-Star on Prefix-Free Languages. Authors. Authors and affiliations.Â We study the star-complement-star operation on prefix-free languages. We get a tight upper bound \(2^{n-3}+2\) for the state complexity of this combined operation on prefix free languages. To prove tightness, we use a binary alphabet. Then we present the results of our computations concerning star-complement-star on binary prefix-free languages. We also show that state complexity of star-complement-star of every unary prefix-free language is one, except for the language \(\{a\}\), where it is two. Keywords.