FORMALIZED MATHEMATICS
2007, Vol. **15**, No. **1**, Pages 11–15
DOI: 10.2478/v10037-007-0002-y

Formal Languages – Concatenation and Closure

Michał Trybulec
Motorola Software Group
Cracow, Poland

Summary. Formal languages are introduced as subsets of the set of all 0-based finite sequences over a given set (the alphabet). Concatenation, the n-th power and closure are defined and their properties are shown. Finally, it is shown that the closure of the alphabet (understood here as the language of words of length 1) equals to the set of all words over that alphabet, and that the alphabet is the minimal set with this property. Notation and terminology were taken from [5] and [13].

MML identifier: `FLANG_1`, version: `7.8.04 4.81.962`

The terminology and notation used here are introduced in the following articles: [10], [4], [11], [8], [9], [2], [14], [3], [1], [6], [12], and [7].

1. Preliminaries

For simplicity, we follow the rules: E is a set, x is a set, A, B, C, D are subsets of E^ω, a, b, c are elements of E^ω, e is an element of E, i, n, n_1, n_2, m are natural numbers, and p, q, r_1, r_2 are real numbers.

Let us consider E, a, b. Then $a \frown b$ is an element of E^ω.

Let us consider E. Then $\langle\rangle_E$ is an element of E^ω.

Let E be a non empty set and let e be an element of E. Then $\langle e\rangle$ is an element of E^ω.

Let us consider E, a. Then $\{a\}$ is a subset of E^ω.

Let us consider E, let f be a function from $\mathbb{N}$ into 2^{E^ω}, and let us consider n. Then $f(n)$ is a subset of E^ω.

One can prove the following propositions:

ISSN 1426–2630

(1) If $\{x\} \not\subseteq X$, then $\{x\}$ misses X.
(2) If $n_1 > 1$ or $n_2 > 1$, then $n_1 + n_2 > 1$.
(3) $n > 0$ iff $n \geq 1$.
(4) If $r_1 + p \leq r_2 + q$ and $p \geq q$, then $r_1 \leq r_2$.
(5) If $n_1 + n \leq n_2 + 1$ and $n > 0$, then $n_1 \leq n_2$.
(6) $n_1 + n_2 = 1$ iff $n_1 = 1$ and $n_2 = 0$ or $n_1 = 0$ and $n_2 = 1$.
(7) $a \frown b = \langle x \rangle$ iff $a = \langle\rangle_E$ and $b = \langle x \rangle$ or $b = \langle\rangle_E$ and $a = \langle x \rangle$.
(8) For all finite 0-sequences p, q such that $a = p \frown q$ holds p is an element of E^ω and q is an element of E^ω.
(9) If $\langle x \rangle$ is an element of E^ω, then $x \in E$.
(10) If $\operatorname{len} b = n + 1$, then there exist c, e such that $\operatorname{len} c = n$ and $b = c \frown \langle e \rangle$.
(11) If $a \frown a = a$, then $a = \emptyset$.

2. Concatenation of Languages

Let us consider E, A, B. The functor $A \frown B$ yields a subset of E^ω and is defined by:

(Def. 1) $x \in A \frown B$ iff there exist a, b such that $a \in A$ and $b \in B$ and $x = a \frown b$.

The following propositions are true:

(12) $A \frown B = \emptyset$ iff $A = \emptyset$ or $B = \emptyset$.
(13) $A \frown \{\langle\rangle_E\} = A$ and $\{\langle\rangle_E\} \frown A = A$.
(14) $A \frown B = \{\langle\rangle_E\}$ iff $A = \{\langle\rangle_E\}$ and $B = \{\langle\rangle_E\}$.
(15) $\langle\rangle_E \in A \frown B$ iff $\langle\rangle_E \in A$ and $\langle\rangle_E \in B$.
(16) If $\langle\rangle_E \in B$, then $A \subseteq A \frown B$ and $A \subseteq B \frown A$.
(17) If $A \subseteq C$ and $B \subseteq D$, then $A \frown B \subseteq C \frown D$.
(18) $(A \frown B) \frown C = A \frown (B \frown C)$.
(19) $A \frown (B \cap C) \subseteq (A \frown B) \cap (A \frown C)$ and $(B \cap C) \frown A \subseteq (B \frown A) \cap (C \frown A)$.
(20) $A \frown B \cup A \frown C = A \frown (B \cup C)$ and $B \frown A \cup C \frown A = (B \cup C) \frown A$.
(21) $A \frown B \setminus A \frown C \subseteq A \frown (B \setminus C)$ and $B \frown A \setminus C \frown A \subseteq (B \setminus C) \frown A$.
(22) $A \frown B \dot{-} A \frown C \subseteq A \frown (B \dot{-} C)$ and $B \frown A \dot{-} C \frown A \subseteq (B \dot{-} C) \frown A$.

3. n-th Power of a Language

Let us consider E, A, n. The functor A^n yields a subset of E^ω and is defined by:

(Def. 2) There exists a function c_1 from $\mathbb{N}$ into 2^{E^ω} such that $A^n = c_1(n)$ and $c_1(0) = \{\langle\rangle_E\}$ and for every i holds $c_1(i + 1) = c_1(i) \frown A$.

Next we state a number of propositions:

(23) $A^{n+1} = (A^n) \frown A$.

(24) $A^0 = \{\langle\rangle_E\}$.

(25) $A^1 = A$.

(26) $A^2 = A \frown A$.

(27) If $n \geq 1$, then $(\emptyset_{E^\omega})^n = \emptyset$.

(28) $\{\langle\rangle_E\}^n = \{\langle\rangle_E\}$.

(29) $A^n = \{\langle\rangle_E\}$ iff $n = 0$ or $A = \{\langle\rangle_E\}$.

(30) If $\langle\rangle_E \in A$, then $\langle\rangle_E \in A^n$.

(31) $(A^n) \frown A = A \frown A^n$.

(32) $A^{m+n} = (A^m) \frown A^n$.

(33) $(A^m)^n = A^{m \cdot n}$.

(34) If $\langle\rangle_E \in A$ and $n > 0$, then $A \subseteq A^n$.

(35) If $\langle\rangle_E \in A$ and $n > 0$ and $m > n$, then $A^n \subseteq A^m$.

(36) If $A \subseteq B$, then $A^n \subseteq B^n$.

(37) $A^n \cup B^n \subseteq (A \cup B)^n$.

(38) $(A \cap B)^n \subseteq A^n \cap B^n$.

(39) If $a \in C^m$ and $b \in C^n$, then $a \frown b \in C^{m+n}$.

4. Closure of a Language

Let us consider E, A. The functor A^* yielding a subset of E^ω is defined as follows:

(Def. 3) $A^* = \bigcup\{B : \bigvee_n \ B = A^n\}$.

The following propositions are true:

(40) $x \in A^*$ iff there exists n such that $x \in A^n$.

(41) $A^n \subseteq A^*$.

(42) If $x \in A$, then $x \in A^*$.

(43) $A \subseteq A^*$.

(44) $A \frown A \subseteq A^*$.

(45) If $a \in C^*$ and $b \in C^*$, then $a \frown b \in C^*$.

(46) If $A \subseteq C^*$ and $B \subseteq C^*$, then $A \frown B \subseteq C^*$.

(47) $A^* = \{\langle\rangle_E\}$ iff $A = \emptyset$ or $A = \{\langle\rangle_E\}$.

(48) $\langle\rangle_E \in A^*$.

(49) If $A^* = \{x\}$, then $x = \langle\rangle_E$.

(50) If $x \in A^{m+1}$, then $x \in (A^*) \frown A$ and $x \in A \frown A^*$.

(51) If $x \in (A^*) \frown A$ or $x \in A \frown A^*$, then $x \in A^*$.

(52) If $\langle\rangle_E \in A$, then $A^* = (A^*) \frown A$ and $A^* = A \frown A^*$.

(53) If $\langle\rangle_E \in A$, then $A^* = (A^*) \frown A^n$ and $A^* = (A^n) \frown A^*$.

(54) $A^* = \{\langle\rangle_E\} \cup A \frown A^*$ and $A^* = \{\langle\rangle_E\} \cup (A^*) \frown A$.

(55) $A \frown A^* = (A^*) \frown A$.

(56) $(A^n) \frown A^* = (A^*) \frown A^n$.

(57) If $A \subseteq B^*$, then $A^n \subseteq B^*$.

(58) If $A \subseteq B^*$, then $A^* \subseteq B^*$.

(59) If $A \subseteq B$, then $A^* \subseteq B^*$.

(60) $(A^*)^* = A^*$.

(61) $(A^*) \frown A^* = A^*$.

(62) $(A^n)^* \subseteq A^*$.

(63) $(A^*)^n \subseteq A^*$.

(64) If $n > 0$, then $(A^*)^n = A^*$.

(65) If $A \subseteq B^*$, then $B^* = (B \cup A)^*$.

(66) If $a \in A^*$, then $A^* = (A \cup \{a\})^*$.

(67) $A^* = (A \setminus \{\langle\rangle_E\})^*$.

(68) $A^* \cup B^* \subseteq (A \cup B)^*$.

(69) $(A \cap B)^* \subseteq A^* \cap B^*$.

(70) $\langle x\rangle \in A^*$ iff $\langle x\rangle \in A$.

5. Alphabet as a Language

Let us consider E. The functor $\mathrm{Lex}\,E$ yielding a subset of E^ω is defined by:

(Def. 4) $x \in \mathrm{Lex}\,E$ iff there exists e such that $e \in E$ and $x = \langle e\rangle$.

Next we state three propositions:

(71) $a \in (\mathrm{Lex}\,E)^{\mathrm{len}\,a}$.

(72) $(\mathrm{Lex}\,E)^* = E^\omega$.

(73) If $A^* = E^\omega$, then $\mathrm{Lex}\,E \subseteq A$.

References

[1] Grzegorz Bancerek. The ordinal numbers. *Formalized Mathematics*, 1(**1**):91–96, 1990.
[2] Czesław Byliński. Functions and their basic properties. *Formalized Mathematics*, 1(**1**):55–65, 1990.
[3] Czesław Byliński. Functions from a set to a set. *Formalized Mathematics*, 1(**1**):153–164, 1990.
[4] Czesław Byliński. Some basic properties of sets. *Formalized Mathematics*, 1(**1**):47–53, 1990.
[5] John E. Hopcroft and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley Publishing Company, 1979.
[6] Beata Padlewska. Families of sets. *Formalized Mathematics*, 1(**1**):147–152, 1990.
[7] Karol Pąk. The Catalan numbers. Part II. *Formalized Mathematics*, 14(**4**):153–159, 2006.

[8] Andrzej Trybulec. Subsets of complex numbers. *To appear in Formalized Mathematics.*
[9] Andrzej Trybulec. Domains and their Cartesian products. *Formalized Mathematics*, 1(**1**):115–122, 1990.
[10] Andrzej Trybulec. Tarski Grothendieck set theory. *Formalized Mathematics*, 1(**1**):9–11, 1990.
[11] Zinaida Trybulec. Properties of subsets. *Formalized Mathematics*, 1(**1**):67–71, 1990.
[12] Tetsuya Tsunetou, Grzegorz Bancerek, and Yatsuka Nakamura. Zero-based finite sequences. *Formalized Mathematics*, 9(**4**):825–829, 2001.
[13] William M. Waite and Gerhard Goos. *Compiler Construction*. Springer-Verlag New York Inc., 1984.
[14] Edmund Woronowicz. Relations defined on sets. *Formalized Mathematics*, 1(**1**):181–186, 1990.

Received March 9, 2007