Acta Mathematica Academiae Paedagogicae Nyíregyháziensis
22 (2006), 121–125
www.emis.de/journals
ISSN 1786-0091

ON A TEST FOR CODES

J. FALUCSKAI

Abstract. Sets of codewords can be represented by finite automata (FAs) and every FA can be represented by connection matrices or regular expressions. Our goal is to find similar systems like that and to solve one of the systems's problems in another system. Having a set of codewords we have to decide whether there are two or more sequences of codewords which form the same chain of characters of codewords. We have developed an algorithm that solves this problem by using finite automata and their deterministic finite automata.

1. Definitions for Code

Let A be a set, which we call an *alphabet*. A *word* w on the alphabet A is a finite sequence of elements of A

$$w = (a_1, a_2, \dots, a_n), \quad a_i \in A$$

The set of all words on the alphabet A is denoted by A^*. A^* is equipped with an associative operation defined by the concatenation of two sequences

$$(a_1, a_2, \dots, a_n)(b_1, b_2, \dots, b_m) = (a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_m).$$

The associativity allows us to write

$$w = a_1 a_2 \dots a_n$$

instead of $w = (a_1, a_2, \dots, a_n)$, by identifying each element $a \in A$ with the sequence (a). An element $a \in A$ is called a *letter*. The empty sequence is called the *empty word* and is denoted by ε. It is the neutral element for concatenation. The set of nonempty words on A is denoted by A^+.

A code C over A is a subset of A^+. The words of C are called *code words*, the elements of C^* are *messages*. A code C is said to be *uniquely decipherable (UD)* if each message has an unique factorization into codewords, i.e. the equality

$$x_1 x_2 \cdots x_n = y_1 y_2 \cdots y_m,$$

$x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_m \in C$, implies $n = m$ and $x_1 = y_1, \dots, x_n = y_n$.

2. An algorithm for uniquely decipherable codes

Our algorithm is based on the automaton theory. This subject was reviewed in [2], [4], [1], [6], but our approach is different from their aspect. We construct an automaton for the code over A by union of automata of codewords. If codeword $w = x_1 x_2 \dots x_n$ then automaton $\mathcal{A}(w)$ of w is $\mathcal{A}(w) = (q_i, Q_t, Q, A, \delta)$ where q_i is the initial state of $\mathcal{A}(w)$ and Q_t is the set of terminal states. Q is the set of states

2000 *Mathematics Subject Classification.* 94B35, 94A45, 68Q45.
Key words and phrases. Uniquely decipherable codes, automata, length-variable codes.

and $Q_t = \{q_i\}$; $\quad q_i \in Q$. $card(Q) = length(w)$ since the rules of automaton $\mathcal{A}(w)$ are the following:

$$\begin{array}{lcl}
\delta(q_i, x_1) & = & q_{x_1} \\
\delta(q_{x_1}, x_2) & = & q_{x_1 x_2} \\
 & \vdots & \\
\delta(q_{x_1 x_2 \ldots x_{n-2}}, x_{n-1}) & = & q_{x_1 x_2 \ldots x_{n-2} x_{n-1}} \\
\delta(q_{x_1 x_2 \ldots x_{n-1}}, x_n) & = & q_i
\end{array}$$

thus $\mathcal{A}(w)$ can recognize w^*. Figure 1 shows the automaton of the codeword 0100.

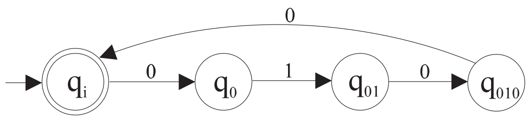

FIGURE 1. Automaton $\mathcal{A}(0100)$

If w_1 is a prefix part of w_2 then their automata will have common states, more exactly $Q^{w_1} \subset Q^{w_2}$. This property occurs in figure 2.

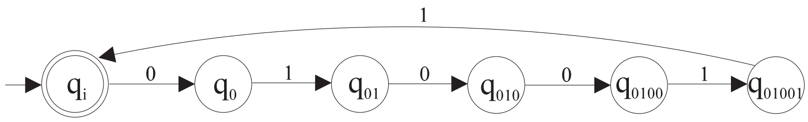

FIGURE 2. Automaton $\mathcal{A}(010011)$

Furthermore $q_i^{w_1} = q_i^{w_2}$ (so $Q_t^{w_1} = Q_t^{w_2}$) and $A^{w_1} = A^{w_2}$. Since w_1 is prefix part of w_2, we can use notation

$$w_1 = x_1 x_2 \ldots x_n; \quad w_2 = x_1 x_2 \ldots x_n x_{n+1} \ldots x_m$$

For rules we get

$\mathcal{A}(w_1)$

$$\begin{array}{lcl}
\delta(q_i, x_1) & = & q_{x_1} \\
\delta(q_{x_1}, x_2) & = & q_{x_1 x_2} \\
 & \vdots & \\
\delta(q_{x_1 x_2 \ldots x_{n-2}}, x_{n-1}) & = & q_{x_1 x_2 \ldots x_{n-2} x_{n-1}} \\
\delta(q_{x_1 x_2 \ldots x_{n-1}}, x_n) & = & q_i
\end{array}$$

$\mathcal{A}(w_2)$

$$\begin{array}{lcl}
\delta(q_i, x_1) & = & q_{x_1} \\
\delta(q_{x_1}, x_2) & = & q_{x_1 x_2} \\
 & \vdots & \\
\delta(q_{x_1 x_2 \ldots x_{n-2}}, x_{n-1}) & = & q_{x_1 x_2 \ldots x_{n-2} x_{n-1}} \\
\delta(q_{x_1 x_2 \ldots x_{n-1}}, x_n) & = & q_{x_1 x_2 \ldots x_{n-1} x_n} \\
 & \vdots & \\
\delta(q_{x_1 x_2 \ldots x_{m-2}}, x_{m-1}) & = & q_{x_1 x_2 \ldots x_{m-2} x_{m-1}} \\
\delta(q_{x_1 x_2 \ldots x_{m-1}}, x_m) & = & q_i
\end{array}$$

Consequently

$$(\delta^{w_1} \setminus \{\delta(q_{x_1 x_2 \ldots x_{m-1}}, x_m) = q_i\}) \subset \delta^{w_2},$$

thus

$$\delta^{w_1} \cup \delta^{w_2} = \delta^{w_2} \cup \{\delta(q_{x_1 x_2 \ldots x_{m-1}}, x_m) = q_i\}.$$

Let

$$\mathcal{A}(w_1, w_2) = (q_i, Q_t = \{q_i\}, Q = Q^{w_1} \cup Q^{w_2}, A, \delta = \delta^{w_1} \cup \delta^{w_2}).$$

Since

$$\begin{array}{rcll} \delta(q_{x_1 x_2 \dots x_{n-1}}, x_n) & = & q_i & \in \delta^{w_1} \\ \delta(q_{x_1 x_2 \dots x_{n-1}}, x_n) & = & q_{x_1 x_2 \dots x_{n-1} x_n} & \in \delta^{w_2}, \end{array}$$

thus $\mathcal{A}(w_1, w_2)$ is non deterministic because the left sides of the rules are equal. $\mathcal{A}(w_1, w_2)$ accepts $\{w_1, w_2\}^*$. Considering the not uniquely decipherable strings, we receive that some codewords of different factoring are prefix, namely if $u_1 \dots u_n = w_1 \dots w_m$, then for all

$$i < j \quad u_i = w_i \text{ holds, but } u_j \neq w_j \text{ where } 1 \leq j \leq \min\{n, m\}.$$

If we join the automata of codewords, then we get the automaton $\mathcal{A}(w_1, \dots, w_n)$ of code $C = \{w_1, \dots, w_n\}$. We can use notation $\mathcal{A}(C)$, too. So

$$\mathcal{A}(C) = (q_i, Q_t = \{q_i\}, Q = Q^{w_1} \cup \dots \cup Q^{w_n}, A, \delta = \delta^{w_1} \cup \dots \cup \delta^{w_n}).$$

Obviously $\mathcal{A}(C)$ accepts C^*. An automaton is non deterministic if there is more then one rule for the same pair of state and symbol.

Theorem 1. *If the automaton $\mathcal{A}(C)$ is deterministic, then the code is uniquely decipherable.*

Proof. If the automaton of the code is deterministic, then the code is prefix (free). The prefix codes are uniquely decipherable. □

Remark 1. There are non prefix codes which are uniquely decipherable, for example $C_1 = \{0100, 010011\}$. Hence if the automaton $\mathcal{A}(C)$ is non deterministic, then the code could be uniquely decipherable. We demonstrate the graphical presentation of $\mathcal{A}(C_1)$ in Figure 3.

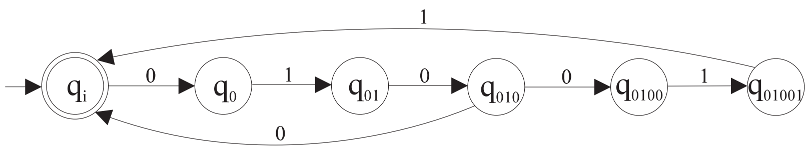

FIGURE 3. Automaton $\mathcal{A}(0100, 010011)$

The automaton is nondeterministic by reason of

$$\begin{array}{rcl} \delta(q_{010}, 0) & = & q_i \\ \delta(q_{010}, 0) & = & q_{0100}, \end{array}$$

but the code is uniquely decipherable. Of course there exist non uniquely decipherable codes with non deterministic automaton, too.

We show a stricter condition of non uniquely decipherability than *Theorem 1.* The construction is based on the well known relationship between deterministic and non deterministic automata. For every non deterministic automaton $\mathcal{A}$ there exists an equivalent deterministic automaton $\mathcal{A}_D$.

If a string S is decipherable on code C then $\mathcal{A}(C)$ accepts S, namely $\mathcal{A}(C)$ reads it and stays in q_i state. If S is not uniquely decipherable then we can follow different paths during reading. We join these different paths by the equivalent deterministic automaton. Formally

$$\overbrace{\underbrace{x_1 \dots x_{|w_{i_1}|}}_{w_{i_1}} \dots x_{|w_{j_1}|}}^{w_{j_1}} \dots \overbrace{\dots\dots\dots \underbrace{\dots x_{|S|}}_{w_{i_m}}}^{w_{j_n}}$$

$$\begin{array}{lcl}
\delta(q_i, x_1) & = & q_{x_1} \\
\delta(q_{x_1}, x_2) & = & q_{x_1 x_2} \\
 & \vdots & \\
\delta(q_{x_1 x_2 \ldots x_{|w_{i_1}|-1}}, x_{|w_{i_1}|}) & = & \{q_{x_1 x_2 \ldots x_{|w_{i_1}|}}, q_i\} \\
\delta(\{q_{x_1 x_2 \ldots x_{|w_{i_1}|}}, q_i\}, x_{|w_{i_1}|+1}) & = & \{q_{x_1 x_2 \ldots x_{|w_{i_1}|+1}}, q_{x_{|w_{i_1}|+1}}\} \\
 & \vdots & \\
\delta(\{q_{x_1 x_2 \ldots x_{|w_{j_1}|-1}}, q_{x_{|w_{i_1}|+1} \ldots x_{|w_{j_1}|-1}}\}, x_{|w_{j_1}|}) & = & \{q_i, q_{x_{|w_{i_1}|+1} \ldots x_{|w_{j_1}|}}\} \\
 & \vdots & \\
\delta(\{q_{x_{|w_{j_n}|+1} \ldots x_{|S|-1}}, q_{x_{|w_{i_m}|+1} \ldots x_{|S|-1}}\}, x_{|S|}) & = & \{q_i, q_i\} = q_i
\end{array}$$

Thus two (or more) factorizations of a string will end by using two (or more) rules with right side q_i.

Theorem 2. *A code is uniquely decipherable if and only if at the most one state is equal to q_i in right side of any rule of $\mathcal{A}_D(C)$, namely for all*

$$\delta(\{q_{i_1}, \ldots, q_{i_n}\}, x) = \{q_{j_1}, \ldots, q_{j_m}\} \in \mathcal{A}_D(C) : \nexists \quad l, k : q_{j_l} = q_{j_k} = q_i$$

Proof. The proof is indirect. If there exists

$$\delta(\{q_{i_1}, \ldots, q_{i_n}\}, x) = \{q_{j_1}, \ldots, q_{j_m}\} \in \mathcal{A}_D(C) : \exists \quad l, k : q_{j_l} = q_{j_k} = q_i$$

then there is a string with at least two different factorizations which is accepted by the automaton. Consequently the code is not uniquely decipherable. ☐

Example 1. Let $C = \{010, 1101, 10, 11\}$. Thus we get $\mathcal{A}(010, 1101, 10, 11)$ in Figure 4. ($q_i = S$). Let us construct $\mathcal{A}_D(010, 1101, 10, 11)$. The result is given in Figure 5. We can see that

$$\delta(\{B, C\}, 0) = \{S, S\} = \{S\}$$

so the code is not uniquely decipherable.

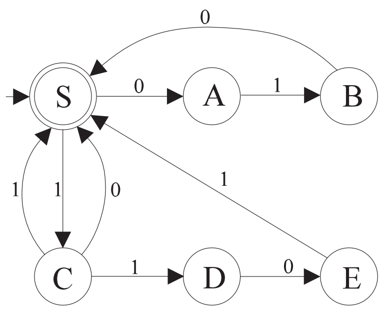

FIGURE 4. Automaton $\mathcal{A}(010, 1101, 10, 11)$

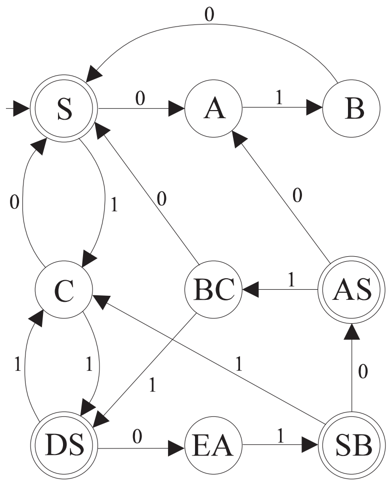

FIGURE 5. Automaton $\mathcal{A}_D(010, 1101, 10, 11)$

REFERENCES

[1] X. Augros and I. Litovsky. Algorithms to test rational ω code. In *Mathematical Foundations of Informatics'99 Conference*, Hanoi, 1999.

[2] J. Berstel and D. Perrin. *Theory of codes*, volume 117 of *Pure and Applied Mathematics*. Academic Press Inc., Orlando, FL, 1985.

[3] F. Burderi and A. Restivo. Varieties of codes and Kraft inequality. In *STACS 2005*, volume 3404 of *Lecture Notes in Comput. Sci.*, pages 545–556. Springer, Berlin, 2005.

[4] R. König. Lectures on codes, 1994. Internal Reports of the IMMD I.

[5] A. A. Sardinas and C. W. Patterson. A necessary and sufficient condition for the unique decomposition of coded messages. In *IRE Internat. Conv. Rec.*, volume 8:104, 1953.

[6] K. Tsuji. An automaton for deciding whether a given set of words is a code. *Sūrikaisekikenkyūsho Kōkyūroku*, 1222:123–127, 2001.

Received September 15, 2005.

INSTITUTE OF MATHEMATICS AND COMPUTER SCIENCE,
COLLEGE OF NYÍREGYHÁZA,
4400 NYÍREGYHÁZA, SÓSTÓI ÚT 31/B,
HUNGARY