Literature

W.I.Löwenstein, Kodierungstheorie. In: *Diskrete Mathematik und mathematische Fragen der Kybernetik*, Herausg.: S.W.Jablonski/O.B.Lupanov, Akademie-Verlag, 1980.

A.Salomaa, *Jewels of Formal Language Theory*. Comp. Sci. Press, 1981.

H.J.Shyr, *Free Monoids and Languages*. Hon Min Book Co., Taichung, Taiwan, 1991.

J. Duske/H.Jürgensen, *Kodierungstheorie*. BI-Taschenb., Mannheim, 1977.

T. Grams, *Codierungsverfahren*. BI-Taschenbuch, Mannheim, 1986.

P. Sweeney, *Codierung zur Fehlererkennung und Fehlerkorrektur*. Hanser-Verlag, 1992.

W.W.Peterson/E.J.Weldon, *Error-Correcting Codes.* MIT Press, Cambridge, 1972.

J.Berstel/D.Perrin, *Theory of Codes*. Academic Press, 1985.

Some Sets

$$
\begin{aligned}
C_0 &= \{a, ba, ab\}, \\
C_1 &= \{a, bb, aab, bab\}, \\
C_2 &= \{aa, bb, aba, baa\}, \\
C_3 &= \{aaa, aba, bab, bbb\}, \\
C_4 &= \{a, ab, bb\}
\end{aligned}
$$

Code – Definition

Definition:

A bijective function $\phi : A \to C$ is called a coding of the set A by the non-empty language C over the alphabet X, if the homomorphic extension of ϕ to A^* is an injective function from A^* into X^*.

A non-empty language C (over X) is called a code, if C is the range of some coding.

Code – Characterisation

Theorem: A non-empty language C is a code if and only if, for any

$$x_{i_1}, x_{i_2}, \ldots, x_{i_n}, x_{j_1}, x_{j_2}, \ldots, x_{j_m} \in C,\ n \geq 1, m \geq 1,$$

the equality $x_{i_1}x_{i_2}\ldots x_{i_n} = x_{j_1}x_{j_2}\ldots x_{j_m}$ implies $x_{i_1} = x_{j_1}$.

Theorem: A language C is a code if and only if, for any

$$x_{i_1}, x_{i_2}, \ldots, x_{i_n}, x_{j_1}, x_{j_2}, \ldots, x_{j_m} \in C,\ n \geq 1, m \geq 1,$$

the equality $x_{i_1}x_{i_2}\ldots x_{i_n} = x_{j_1}x_{j_2}\ldots x_{j_m}$ implies

$$n = m \quad \text{and} \quad x_{i_t} = x_{j_t} \text{ for } 1 \leq t \leq n.$$

Strong Code

Definition: A code C is called a strong code,
if for any $x_{i_k} \in C$ and $x_{j_k} \in C$, $k \geq 1$, and any $n \geq 1$ such that
$x_{i_1} x_{i_2} \ldots x_{i_n}$ is a prefix of $x_{j_1} x_{j_2} \ldots x_{j_n}$ or
$x_{j_1} x_{j_2} \ldots x_{j_n}$ is a prefix of $x_{i_1} x_{i_2} \ldots x_{i_n}$,
the equality $x_{i_1} = x_{j_1}$ holds.

Remark: A code C is a strong code if and only if,
for any $x_{i_k} \in C$ and $x_{j_k} \in C$, $k \geq 1$, and any $n \geq 1$ such that
$x_{i_1} x_{i_2} \ldots x_{i_n}$ is a prefix of $x_{j_1} x_{j_2} \ldots x_{j_n}$ or
$x_{j_1} x_{j_2} \ldots x_{j_n}$ is a prefix of $x_{i_1} x_{i_2} \ldots x_{i_n}$,
the equalities $x_{i_k} = x_{j_k}$ hold for $k \geq 1$.

Special Codes

Definition:
A non-empty language C is called a prefix code, if no word of C is a prefix of another different word of C.

Definition: Let $n \geq 1$ be a natural number. A subset C of X^n is called a block code of length n over X.

Theorem:
For any code C and any natural number $k \geq 1$, C^k is a code, too.

Decoding

Definition: A Mealy automaton is a 6-tuple $\mathcal{A} = (X, Y, Z, f, g, z_0)$ where
– X, Y, Z are alphabets (finite non-empty sets)
– $f : Z \times X \to Z$ and $g : Z \times X \to Y^*$ are functions, and
– z_0 is an element of Z.

f and g are extended to $Z \times X^*$ by

$$f^*(z, \lambda) = z, \qquad g^*(z, \lambda) = \lambda,$$
$$f^*(z, wa) = f(f^*(z, w), a), \qquad g^*(z, wa) = g^*(z, w)g(f^*(z, w), a)$$

for $w \in X^*$, $a \in X$

Theorem:
There is an algorithm which, for any strong coding $\phi : A \to C \subseteq X^+$ and any word $x \in X^+$, computes in linear time $\phi^{-1}(x)$ or detects in linear time that $\phi^{-1}(x)$ is not defined.