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

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

Code – Definition

Definition:

A bijective function $\varphi : A \to C$ is called a coding of the set A by the non-empty language C over an 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
$f^*(z, \lambda) = z$, $g^*(z, \lambda) = \lambda$,
$f^*(z, wa) = f(f^*(z, w), a)$, $g^*(z, wa) = g^*(z, w)g(f^*(z, w), a)$
for $w \in X^*$, $a \in X$

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

Product Independent Sets

Definition:
A language L is called product independent, if no word of L can be represented as the product of at least two words from L.

Theorem: Let C be a product independent set over X. Then, C is exactly then a code, if, for any word $w \in X^*$,

$$wC^* \cap C^* \neq \emptyset \text{ and } C^*w \cap C^* \neq \emptyset \quad \text{imply} \quad w \in C^*.$$

Decidability of the Code Property

Theorem: Let $C = \{x, y\}$ be a set with two non-empty words over X. Then, C is exactly then a code, if $xy \neq yx$.

$$K_0(C) = C,$$
$$K_{i+1}(C) = \{w \in X^+ \mid yw = x \text{ or } xw = y \text{ for certain } x \in C, y \in K_i(C)\}.$$

Theorem: A non-empty language C over X is exactly then a code, if $K_i(C) \cap C = \emptyset$ for $i \geq 1$.

Theorem: A code C over X is exactly then a strong code, if $K_n(C) = \emptyset$ for $n \geq \#(C)(\max\{|c| \mid c \in C\} - 1) + 1$.

Theorem: There is an algorithm which decides, for every finite language C over a finite alphabet X, whether C is a (strong) code.

Two Lemmas

Lemma: For every code C, every $n \geq 0$, and every $w \in K_n(C)$, we have $w \in \mathit{Suff}(C)$.

Lemma: A word v_n is an element of $K_n(C)$ $(n \geq 1)$ if and only if, for every $i < n$, there are words $v_i \in K_i(C)$ and $x_{i_1}, x_{i_2}, \ldots, x_{i_k}, x_{j_1}, x_{j_2}, \ldots, x_{j_l} \in C$ with $k + l = n - i$ such that either

$$v_i x_{i_1} x_{i_2} \ldots x_{i_k} v_n = x_{j_1} x_{j_2} \ldots x_{j_l} \quad \text{with} \quad |v_n| < |x_{j_l}|$$

or

$$v_i x_{i_1} x_{i_2} \ldots x_{i_k} = x_{j_1} x_{j_2} \ldots x_{j_l} v_n \quad \text{with} \quad |v_n| < |x_{i_k}| \text{ for } k \neq 0$$

holds.

Code Indicator I

Definition:

Let X be an alphabet of the cardinality $n \geq 2$. The code indicator $ci_n(w)$ of a word $w \in X^*$ is defined by

$$ci_n(w) = n^{-|w|}.$$

For a language L with n_L letters, we set

$$ci(L) = \sum_{w \in L} ci_{n_L}(w).$$

Code Indicator II

Theorem:
Let L_1 and L_2 be two languages over a minimal alphabet X consisting of n letters. Then

$$ci(L_1 \cdot L_2) \leq ci(L_1) \cdot ci(L_2),$$

and the equality holds if and only if, for any four words $w_1, w_2 \in L_1$ and $w_3, w_4 \in L_4$, the equality $w_1 w_3 = w_2 w_4$ implies $w_1 = w_2$.

Theorem:
For each code C, we have $ci(C) \leq 1$.

Code Indicator III

Theorem:
Let $n \geq 2$ and $l_1, l_2, \ldots, l_m$ be natural positive numbers that satisfy

$$\sum_{i=1}^{m} n^{-l_i} \leq 1.$$

Then a code (prefix code)

$$C = \{c_0, c_1, \ldots, c_{m-1}\}$$

over an alphabet X with n elements exists where

$$|c_{i-1}| = l_i \quad \text{for} \quad 1 \leq i \leq m.$$

Maximal Codes

Definition:
A code C is called maximal, if for each word $w \notin C$, the set $C \cup \{w\}$ is not a code.

Theorem:
A code C with $ci(C) = 1$ is a maximal code.

Theorem:
A finite code C is maximal if and only if $ci(C) = 1$.