Journal of Universal Computer Science, vol. 6, no. 1 (2000), 130-135
submitted: 15/6/99, accepted: 2/11/99, appeared: 28/1/00 © Springer Pub. Co.

Free–Extendible Prefix–Free Sets and an Extension of the Kraft–Chaitin Theorem[1]

Cristian Grozea
(Faculty of Mathematics
Bucharest University
Str. Academiei 14, R-70109 Bucharest, Romania
Email: chrisg@phobos.cs.unibuc.ro)

Abstract: First, the dual set of a finite prefix–free set is defined. Using this notion we describe equivalent conditions for a finite prefix–free set to be indefinitely extendible. This lead to a simple proof for the Kraft–Chaitin Theorem. Finally, we discuss the influence of the alphabet size on the indefinite extensibility property.

Key Words: Prefix–free set, Kraft's inequality.
Category: F2.2

1 Introduction

Continuing the study of prefix–free extendible sets in [5, 6] we obtain a necessary and sufficient condition for for a finite prefix–free set to be indefinitely extendible. As a consequence, a simple argument proving the Kraft–Chaitin Theorem is obtained. We will also study conditions under which the extensibility property is invariant under the change of the size of the alphabet.

2 Notation

Let $\mathbf{N}=\{0,1,2,3,\dots\}$ be the set of the non-negative integers. Fix the alphabet $\Sigma = \{a_1, a_2, \dots, a_Q\}$ (where $Q \geq 2$) and denote by Σ^* the set of all words over the alphabet Σ. We shall denote by $\#(S)$ the size of the finite set S, by $\mid x \mid$ the length of the word $x \in \Sigma^*$ and by $x \subset y$ the fact that x is a prefix of y (for $x, y \in \Sigma^*$). A set $M \subset \Sigma^*$ is called prefix–free if there are no words x and y, $x \in M$ and $y \in M$, with $x \neq y$ and $x \subset y$.

We consider the μ "measure", defined by $\mu(M) = \sum_{x \in M} Q^{-|x|}$. *For any prefix–free set $M \subset \Sigma^*$ we have $\mu(M) \leq 1$* (Kraft inequality).

For more details see more in [1].

3 The Dual Set

Let $M \subset \Sigma^*$ be finite and prefix–free. An *extension* of M is a word x, $x \notin M$, such that $M \cup \{x\}$ is still prefix–free. An *extension root* of M is a minimal extension for M, i.e., x is an extension of M, but no proper prefix of x is an

[1] C. S. Calude and G. Ştefănescu (eds.). *Automata, Logic, and Computability. Special issue dedicated to Professor Sergiu Rudeanu Festschrift.*

extension of M (if $y \subset x$ then $m \subset y$ or $y \subset m$, for some $m \in M$). Denote by $D(M)$ the set of all extension roots of M.

Take $\Sigma = \{a, b, c\}$. Here are some examples:

- $D(\{ab, ac\}) = \{aa, b, c\}$.
- $D(\{a\}) = \{b, c\}$.
- $D(\{abc\}) = \{aba, abb, aa, ac, b, c\}$.
- $D(\emptyset) = \{\lambda\}$.
- $D(\{\lambda\}) = D(\Sigma) = D(\Sigma^n) = \emptyset$, for all $n > 0$.

Definition 3.1 *Let $M \subset \Sigma^*$ be a prefix–free finite set. The (prefix) dual of M is the set $D(M)$ of the extension roots of M.*

If we consider the tree representation of the prefix–free set M, its dual $D(M)$ has as elements the words corresponding to the extra leafs that need to be added in order to transform the tree into a *complete tree*, a tree such that each internal node has exactly Q children.

Proposition 3.2 *The following statements are true:*

1. $D(\emptyset) = \{\lambda\}$.
2. $M \cap D(M) = \emptyset$.
3. *The dual $D(M)$ is a finite prefix–free set. (More exactly, $\#D(M) < Q \cdot \#M \cdot max_{x \in M} \mid x \mid +1$.)*
4. *Any word that can be used to extend M in a prefix–free manner has a prefix in $D(M)$.*
5. *For each $x \in D(M)$, $D(M \cup \{x\}) = D(M) \setminus \{x\}$.*
6. $\mu(M) + \mu(D(M)) = 1$.

Proof. The first two properties follows directly from definitions. For the third property, the tree representation of the finite prefix–free set M and the definition of $D(M)$ show that $D(M)$ has no element longer than the longest element of M, so it is finite.

As a direct consequence of its construction $D(M)$ is prefix–free: if $w \in D(M), w \neq \lambda$, then there is another word $\alpha \in M$ such that $w[1.. \mid w \mid -1]$ is a prefix of α. This leads to the formula $\#D(M) < Q \cdot \#M \cdot max_{x \in M} \mid x \mid +1$.

Here is a simple *algorithm for computing* $D(M)$:

0. Put $D = \emptyset$;
1. If $M = \emptyset$, then $D = D \cup \{\lambda\}$; display D; stop.
2. If $M = \{\lambda\}$, then display D; stop.
3. Let w be one of the words with maximum length in M (it does not matter which one we pick up). We know already (see step 2) that $w \neq \lambda$.

   Let x be the root of the w, obtained from w by deleting its last symbol. So x is a prefix of w and its length is $\mid x \mid = \mid w \mid -1$.

   Next we extend D

$$D = (D \setminus \{w\}) \cup \{xa_i \mid i \in 1, \ldots, Q, xa_i \notin M\}$$

`and transform` M

$$M = (M \setminus \{xa_i \mid i = 1, \ldots, Q\}) \cup \{x\},$$

`and go to step 1.`

The next (fourth) property follows again from the definitions of the extension word and extension roots.

For the fifth property, it is enough to observe that the extension words for $M \cup \{x\}$ are the same as for the set M, excepting those having the word x as prefix. For this reason, the extension roots of the set $M \cup \{x\}$ are the same as the extension roots for the set M, excepting the word x.

For the last (sixth) property of $D(M)$, notice that the measure $\mu(S) = 1$ if the associated tree of the finite set S is complete (this is easy to prove by induction using a node collapsing argument).

Next notice that if S and T are two finite, prefix–free, disjoint sets of words and $S \cup T$ is prefix–free, then $\mu(S \cup T) = \mu(S) + \mu(T)$. It follows that M and $D(M)$ have complementary measures with respect to μ:

$$\mu(M) + \mu(D(M)) = \mu(M \cup D(M)) = 1.$$

□

4 Free–Extendible Prefix–Free Sets

Definition 4.1 *A prefix–free set $M \subset \Sigma^*$ is free–extendible if it can be extended in any way that does not conflict to Kraft's inequality: for any word lengths $(\varphi_n)_{n \in \mathbf{I}}$ ($\mathbf{I}$ is either an initial prefix of $\mathbf{N}$ or the whole set $\mathbf{N}$) such that*

$$\sum_{n \in \mathbf{I}} Q^{-\varphi_n} \leq 1 - \sum_{x \in M} Q^{-|x|},$$

we can construct the words $(y_n)_{n \in \mathbf{I}}, y_n \in \Sigma^*$ such that

1. $\mid y_n \mid = \varphi_n$, *for each* $n \in \mathbf{I}$,
2. $\{y_n \mid n \in \mathbf{I}\} \cap M = \emptyset$,
3. $\{y_n \mid n \in \mathbf{I}\} \cup M$ *is prefix–free.*

Definition 4.2 *The profile of a set of words $M \subset \Sigma^*$ is the histogram of the lengths of words in M:*

$$profile(M)(i) = \#\{x \in M \mid\mid x \mid = i\}, \ for \ i \in \mathbf{N}.$$

Definition 4.3 *A set of words M has a thin profile (over Σ) if its profile is bounded by $Q - 1$ (where $Q = \#\Sigma$): $profile(M)(i) \leq Q - 1$, for each $i \in \mathbf{N}$.*

Theorem 4.4 *A finite prefix–free set of words $M \subset \Sigma^*$ is free–extendible if and only if its dual $D(M)$ has a thin profile.*

Proof. Assume M is free–extendible and its (finite) dual $D(M)$ has not a thin profile; let $p = \min\{i \in \mathbf{N} \mid profile(D(M))(i) \geq Q\}$.

We prove now that we may assume that

$$profile(D(M))(i) = 0, \text{ for } i < p. \tag{1}$$

If this is not the case, we can transform M into another set M', still finite, prefix–free and free–extendible, such that the property above holds for M'. Let $i_0 = \min\{i \in \mathbf{N} \mid profile(D(M))(i) > 0\}$. After the transformation,

$$profile(D(M'))(i) = profile(D(M))(i) = 0, \text{ for } i < i_0,$$

$$profile(D(M'))(i_0) = 0,$$

and

$$profile(D(M'))(i) = profile(D(M))(i), \text{ for all } i > i_0.$$

As $\mu(M) + \mu(D(M)) = 1$ (see (3.2)) and

$$\mu(D(M)) = \sum_{i \in \mathbf{N}} profile(D(M))(i) Q^{-i},$$

it follows that

$$profile(D(M))(i_0) Q^{-i_0} \leq 1 - \mu(M).$$

Now put $\varphi_i = i_0$, $i \in \mathbf{I} = \{0, \ldots, profile(D(M))(i_0) - 1\}$; it follows the existence of the words $(y_n)_{n \in \mathbf{I}}$ with properties described in Definition 4.1. Notice that $\{y_n \mid n \in \mathbf{I}\} \subset D(M)$, because otherwise y_n would have proper prefixes in $D(M)$, which is not possible because $\mid y_n \mid = i_0$ and there is no word in $D(M)$ with the length smaller than i_0.

Let $M' = M \cup \{y_n \mid n \in \mathbf{I}\}$; then $D(M') = D(M) \setminus \{y_n \mid n \in \mathbf{I}\}$.

This concludes the description of the transformation of M. This transformation, which may be applied repeatedly, justifies our claim that we may assume that $profile(D(M))(i) = 0$, for $i < p$, and $profile(D(M))(p) >= Q$. ¿From this we deduce that $\mu(D(M)) \geq Q \cdot Q^{-p} = Q^{-(p-1)}$ and therefore $\mu(M) \leq 1 - Q^{-(p-1)}$.

Now, we could ask for a $p - 1$ length word to extend the free–extendible set M, as this request does not conflict with Kraft's inequality: put $\mathbf{I} = \{0\}$, $\varphi_0 = p - 1$ and use again the Definition 4.1; there exist a word $y_0 \in \Sigma^{(p-1)}$ which can be used to extend M.

¿From properties of the dual, M can only be extended through words taken from its dual set $D(M)$, which doesn't contains any word shorter than p.

So we have got a contradiction that proves that $D(M)$ must have thin profile in order that M to be free–extendible.

For the converse implication assume that $D(M)$ has thin profile. We will prove that M can be indefinitely extended as long as the word length requested does not lead to a violation of Kraft's inequality.

First note that if $Q^{-\varphi} + \mu(M) \leq 1$ then $Q^{-\varphi} \leq \mu(D(M))$ because $D(M)$ has a thin profile it contains a word of length φ or shorter.

Then if we choose to extend any word in $D(M)$ with maximal length but still less than φ, with any prefix of appropriate length and denote by M' the resulted set, then $D(M')$ remains with a thin profile.

This assures the possibility to indefinitely serve another request, maintaining after transformation the property that the dual has a thin profile. □

It's nice to notice that the values taken by $profile(D(M))$ in $0, 1, 2, 3, 4, \ldots$ are the digits of the Q-ary expansion of the real number $\mu(D(M))$; this holds true in general for any finite, set with a thin profile, so it holds true for $D(M)$ too.

Lemma 4.5 *The empty set is a free–extendible prefix–free set for any alphabet size $Q \geq 2$.*

Proof. Clearly, $D(\emptyset) = \{\lambda\}$. The profile of the singleton $\{\lambda\}$ is at most $1 \leq Q-1$, for any alphabet size $Q \geq 2$. □

Corollary 4.6 ([4, 1, 3]) {Kraft–Chaitin Theorem} *Let $\varphi : \mathbf{N} \xrightarrow{o} \mathbf{N}$ be a p.r. function having as domain an initial segment of $\mathbf{N}$. The following two statements are equivalent:*

(1) We can effectively construct an injective p.r. function

$$\theta : dom(\varphi) \xrightarrow{o} \Sigma^*$$

such that:

a) for every $n \in dom(\varphi)$, $\mid \theta(n) \mid = \varphi(n)$,

b) $range(\theta)$ is prefix free.

(2) One has: $\Sigma_{i \in dom(\varphi)} Q^{-\varphi(i)} \leq 1$.

Proof. The direct implication follows directly from the Kraft–Chaitin inequality. The reverse implication holds true because $\emptyset$ is free–extendible.

5 Applications

Note that both the notions of "dual set" and "thin profile" depend upon the size of the underlying alphabet (see [2] for a discussion on the influence of the size of the alphabet on the complexity). This motivates the following

Question 1. What happens with some finite, prefix–free and free–extendible set M if we are changing the size of the underlying alphabet?

The following propositions answer the question.

Proposition 5.1 *The free-extensibility property is preserved as the alphabet size is reduced*

(when this alphabet shrinking is possible, i.e., the set M does not cover all of the alphabet symbols Σ).

Proof. If the alphabet is reduced by exactly one symbol, say a, then a is not allowed to occur in any of the words of M, so it necessarily occurs as a $D(M)$ leaf of at least one node on each tree level.

This means that the new profile of the dual set $D(M)$ is at least reduced by one, on each tree level apart from the root:

$$profile(D_{Q-1}(M))(i) \leq profile(D_Q(M))(i) - 1 \leq (Q-1) - 1 = Q - 2,$$

for $i > 0$.

As the root level of $D(M)$ contains always at most one node (the root of the tree), and

$$profile(D_Q(M)) \leq Q - 1,$$

it follows that

$$profile(D_{Q-1}(M)) \leq Q - 2.$$

¿From Theorem 4.4 it follows that M is still free–extendible over the reduced alphabet that contains $Q - 1$ symbols.

It is obvious now that it is possible to iterate this process for removing more than one symbol from the initial alphabet Σ. It suffices to remove one symbol at a time and to proceed step by step; after each such step the set M remaining free–extendible. □

As a generalization of Lemma 4.5 we get

Proposition 5.2 *The only finite sets preserving their free–extensibility property on arbitrary alphabet extensions are those containing at most one internal (non-leaf) node on each of levels of the associated tree.*

Proof. If there are at least two such internal nodes on some tree level, then after each symbol insertion into the alphabet the profile of the $D(M)$ on the next tree level grows at least with 2.

Even if at some moment the relation $profile(D(M)) \leq Q - 1$ holds, it will soon become false (in at most Q steps), because $profile(D(M))$ (on that level containing children of the two internal nodes on the same level) grows at least twice as fast as Q.

This means that adding more than Q symbols to the alphabet Σ makes M loose its free–extendability (which depends on M and Σ).

We shall show that it is enough to have at most one internal node on each level of the tree associated with the set M, to guarantee that M is indefinitely free–extendible, while new symbols are added to the alphabet Σ. Indeed, let us notice that the profile of $D(M)$ grows at most with the same speed as Q (the growth is 0 for the root level, and 1 for any other level). Hence, the relation $profile(D_Q(M)) \leq Q - 1$ will be true indefinitely as Q grows, so using Theorem 4.4, the set M is free–extendible over any alphabet with more than Q elements. □

References

[1] C. Calude. *Information and Randomness. An Algorithmic Perspective,* Springer-Verlag, Berlin, 1994.

[2] C. Calude, C. Câmpeanu. Are binary codings universal?, *Complexity* 1, 15 (1996), 47-50.

[3] C. Calude and C. Grozea. Kraft-Chaitin inequality revisited, *J. Univ. Comput. Science* 2 (1996), 306-310.

[4] G. J. Chaitin. *Information, Randomness and Incompleteness, Papers on Algorithmic Information Theory*, World Scientific, Singapore, 1987. (2nd ed., 1990)

[5] I. Măndoiu. Kraft-Chaitin's theorem for free-extensible codes, *Studii şi Cercetări Matematice* 44 (1992), 497-501. (Romanian)

[6] I. Măndoiu. Optimum extensions of prefix codes, *Information Processing Letters* 66 (1998), 35-40.