Minimal Covers of Formal Languages

Michael Domaratzki[1], Jeffrey Shallit[2⋆], and Sheng Yu[3]

[1] Department of Computing and Information Science, Queen's University
Kingston, Ontario, Canada K7L 3N6
domaratz@cs.queensu.ca
[2] Department of Computer Science, University of Waterloo
Waterloo, Ontario, Canada N2L 3G1
shallit@graceland.uwaterloo.ca
[3] Department of Computer Science, University of Western Ontario
London, Ontario, Canada N6A 3K7
syu@csd.uwo.ca

Abstract. Let L, L' be languages. If $L \subseteq L'$, we say that L' covers L. Let $\mathcal{C}, \mathcal{D}$ be two classes of languages. If $L' \in \mathcal{C}$, we say that L' is a minimal $\mathcal{C}$-cover with respect to $\mathcal{D}$ if whenever $L \subseteq L'' \subseteq L'$ and $L'' \in \mathcal{C}$, we have $L' - L'' \in \mathcal{D}$. In this paper we discuss minimal $\mathcal{C}$-covers with respect to finite languages, when $\mathcal{C}$ is the class of regular languages.

1 Introduction

Let $L, L' \subseteq \Sigma^*$ be languages. If $L \subseteq L'$, then we say L' *covers* L. In this paper, we are interested in studying the case where the covering language L' is

(i) restricted to lie in some language class—in particular, the regular languages (REG)—and
(ii) is *minimal* in some sense.

The motivation for studying minimal covers is that arbitrary languages L may be arbitrarily difficult to recognize. However, a regular cover L' is easy to recognize, and if the regular cover is minimal, then we might hope that the difference between L' and L is not too large. A recognition algorithm based on L' will mistakenly accept some words it shouldn't, but never mistakenly reject a word in L.

One definition of minimal that at first sight seems natural is the following. Let $\mathcal{C}$ be a class of languages. If $L' \in \mathcal{C}$, we might say L' is a minimal $\mathcal{C}$-cover of L if $L'' \in \mathcal{C}$ and $L \subseteq L'' \subseteq L'$ implies $L'' = L'$. If $\mathcal{C}$ is closed under finite modification — which is the case for nearly every interesting class of languages — then under this definition only members of $\mathcal{C}$ would have minimal covers, and every language is minimally covered by itself! Suppose that L has a minimal cover $L' \in \mathcal{C}$. Then either $L = L'$, in which case $L' \in \mathcal{C}$, or $L \neq L'$. In the latter case, choose any $x \in L' - L$, and consider $L'' = L' - \{x\}$. Since $\mathcal{C}$ is closed under finite modification, $L'' \in \mathcal{C}$. Now $L \subseteq L'' \subseteq L'$, so L' was not minimal, a contradiction.

Thus it is clear we need to seek an alternative definition of minimal cover. To do so we introduce a second language class, $\mathcal{D}$.

⋆ Research supported in part by a grant from NSERC.

W. Kuich, G. Rozenberg, and A. Salomaa (Eds.): DLT 2001, LNCS 2295, pp. 319–329, 2002.

Definition 1. *We say that L' is a minimal $\mathcal{C}$-cover of L with respect to $\mathcal{D}$ if the following conditions hold:*

(i) $L' \in \mathcal{C}$, $L \subseteq L'$ *and*
(ii) *for all languages* $L'' \in \mathcal{C}$ *with* $L \subseteq L'' \subseteq L'$, *we have* $L' - L'' \in \mathcal{D}$.

The class of languages L having a minimal $\mathcal{C}$-cover with respect to $\mathcal{D}$ is denoted by $MC(\mathcal{C}, \mathcal{D})$.

An alternative characterization of minimal covers is possible in some cases, as follows.

Proposition 1. *Suppose $\mathcal{C}$ is a class of languages that is closed under intersection and complement. Suppose $L \subseteq L'$ and $L' \in \mathcal{C}$. Then L' is a minimal $\mathcal{C}$-cover of L with respect to $\mathcal{D}$ if and only if every subset $S \subseteq L' - L$ with $S \in \mathcal{C}$ satisfies $S \in \mathcal{D}$.*

Proof. Suppose L' is a minimal $\mathcal{C}$-cover with respect to $\mathcal{D}$. Let $S \subseteq L' - L$ and $S \in \mathcal{C}$. Now let $L'' := L' - S$. Since $\mathcal{C}$ is closed under intersection and complement, we have $L'' \in \mathcal{C}$. But then $S = L' - L'' \in \mathcal{D}$.

On the other hand, suppose every subset $S \subseteq L' - L$ with $S \in \mathcal{C}$ satisfies $S \in \mathcal{D}$. Let $L'' \in \mathcal{C}$ be such that $L \subseteq L'' \subseteq L'$. Define $S := L' - L''$; then by the assumed closure properties $S \in \mathcal{C}$. Since $S \subseteq L' - L$, it follows that $L' - L'' = S \in \mathcal{D}$. ■

If a language L is infinite and no infinite subset of L lies in the class $\mathcal{C}$, it is said to be $\mathcal{C}$-*immune*. The terminology was apparently introduced by Post [11], who proved among other things that if L is an infinite recursively enumerable set, then L is not `RECURSIVE`-immune. For other works on immunity, see, for example, [1, p. 13] and [13, p. 107].

In this paper, our main focus is when $\mathcal{C} = \texttt{REG}$ and $\mathcal{D} = \texttt{FINITE}$. In this case, Proposition 1 can be rephrased as follows:

Proposition 2. *Let L be a non-regular language, L' be a regular language, and $L \subseteq L'$. Then L' is a minimal regular cover of L with respect to finite languages iff $L' - L$ is* `REG`*-immune.*

Proof. We have $T := L' - L$ is infinite, for if T were finite, then $L = L' - T$ would be regular, a contradiction. Now use Proposition 1. ■

(Flajolet and Steyaert briefly mentioned `REG`-immunity in a 1974 paper [4].)

We point out that the term "minimal cover" was recently used by Câmpeanu, Sântean, and Yu [2] in a different context.

2 Some Examples

In this section we consider some specific examples of context-free languages and determine if they have minimal regular covers. These examples show that there exist context-free languages that do not have a minimal regular cover, and there exist non-regular context-free languages that do have minimal regular covers.

We recall two theorems of Lyndon and Schützenberger [10]: