COMMENTATIONES MATHEMATICAE UNIVERSITATIS CAROLINAE

17,2 (1976)

RECOGNIZABLE FILTERS AND IDEALS

Václav BENDA, Kamila BENDOVÁ, Praha

Abstract: Necessary and sufficient conditions are obtained for filters, ultrafilters, and ideals over a free monoid to be recognizable by finite branching automata.

Key-words: Filter, ultrafilter, ideal, formal language, recognizable family of languages, finite branching automaton.

AMS: 02F10, 02J05 Ref. Ž.: 2.724

Recognizable families of formal languages were introduced and studied in connection with formalization of certain aspects of state-space problem solving by means of finite branching automata (see [1]). In that formalism languages (sets of strings over a finite alphabet Σ) represent plans of behaviour incorporating branching. In an earlier paper [2] we obtained a series of results concerning recognizable families of languages as well as their interesting subclass, the well-recognizable families (recognizable families with recognizable complements).

In the present paper we focus on a particular problem concerning the relationship between recognizable families of languages on one hand and filters and ideals over the free monoid Σ^* on the other hand. The concept of a filter, and its dual notion of an ideal, are important in various areas of mathematics: filters over Σ^* were discussed in [3] especially in connection with concatenation of families.

Here we shall obtain necessary and sufficient conditions for filters and ideals over Σ^* to be recognizable. We shall also show that a recognizable filter is well-recognizable iff it is an ultrafilter. Thus concepts approached from completely different directions appear surprisingly interrelated.

In the present context an alphabet Σ is an arbitrary finite non-empty set of objects called letters (usually denoted a,b,c...). We denote by Σ^* the set of all finite sequences of letters (the free monoid generated by Σ^* under concatenation). The elements of Σ^* are called strings and usually denoted u,v,w... The unit element in Σ^* is the empty string $\Lambda \in \Sigma^*$. We denote $\Sigma_\Lambda = \Sigma \cup \{\Lambda\}$. For $u \in \Sigma^*$, $lg(u)$ denotes the length of u (the number of occurrences of letters in u). In particular, $lg(\Lambda) = 0$. For $u,v \in \Sigma^*$, $u \leq v \equiv$ $\equiv (\exists w \in \Sigma^*)\,(uw = v)$. $\mathcal{P}(\Sigma^*)$ is the set of all subsets of Σ^*, $\mathcal{L}(\Sigma)$ is the set of all non-empty subsets of Σ^*, elements of $\mathcal{L}(\Sigma)$ are called languages (usually denoted L). Any $X \subseteq \mathcal{L}(\Sigma)$ will be called a family of languages (over Σ). Note that we admit empty family of languages but not families with empty element. We shall use the usual set-theoretical operations, union ($\cup$), intersection ($\cap$) and complement ($\overline{X} =$ $= \{L;\ L \in \mathcal{L}(\Sigma)\ \&\ L \notin X\}$). For $u \in \Sigma^*$ and $L \in \mathcal{L}(\Sigma)$ we define:

1) the derivative of L with respect to u

$$\partial_u L = \{ v;\ v \in \Sigma^* \ \&\ uv \in L \};$$

2) the prefix closure of L

$$\mathrm{Pref}(L) = \{ u;\ (\exists v \in L)\ (u \leq v) \};$$

3) the set of first letters of L

$$\mathrm{Fst}\ (L) = \mathrm{Pref}\ (L) \cap \Sigma\ ;$$

4) $\mathrm{Fst}_\Lambda(L) = \mathrm{Pref}\ (L) \cap \Sigma_\Lambda$.

<u>Definition 1</u>. The derivative of a family X with respect to u is the family

$$\partial_u X = \{ \partial_u L;\ L \in X \} - \{ \emptyset \}.$$

We denote $D(X) = \{ \partial_u X;\ u \in \Sigma^* \}$ and we say that X is finitely derivable if D(X) is finite.

<u>Definition 2</u>. C-closure of a family X is the family $C(X) = \{L;\ (\forall u \in \Sigma^*)\ (\exists L_u \in X)\ [\mathrm{Fst}_\Lambda(\partial_u L) =$

$= \mathrm{Fst}_\Lambda(\partial_u L_u)]\}$.

We say that a family X is self-compatible if C(X) = X.

Recognizable families of languages were originally defined in terms of finite branching automata (hence the attribute "recognizable"). Here we shall need only their structural characterization (see [1]), which we shall use, therefore, as a definition.

<u>Definition 3</u>. A family X is recognizable if X is self-compatible and finitely derivable.

Let us note that, as it is known from classical automata theory, a language L is regular (i.e. recognizable by a classical finite automaton) iff the set $\{\partial_u L;\ u \in \Sigma^*\}$ is finite. The reader unfamiliar with the automata theory may consider this fact as a definition of a regular language. (Note that in the classical automata theory $\emptyset$ is also a regular language.)

For the definition and basic properties of filters, see e.g.[4] IV,8, p. 193-196.

Definition 4. A filter F over Σ^* is a non-empty subset of $\mathcal{P}(\Sigma^*)$ satisfying:

1) $\emptyset \notin F$;

2) if $A, B \in F$ then $A \cap B \in F$;

3) if $A \in F$ and $A \subseteq B$ then $B \in F$.

In this paper we assume Σ to be a fixed alphabet and shall call filters over Σ^* simply filters.

Since $\emptyset \notin F$ every filter is a subset of $\mathcal{L}(\Sigma)$ and we can look at it as a family of languages. For any $L \in \mathcal{L}(\Sigma)$ the family $\{L'; L \subseteq L'\}$ is clearly a filter over Σ^*. Over an infinite set there exist also filters of other types (here e.g. family of all languages with finite complements).

Definition 5. A filter of the type $\{L'; L \subseteq L'\}$ is called principal and will be written F_L.

It is easy to show that a filter F is principal iff $\bigcap F \in F$.

Definition 6. A filter F is called an ultrafilter if F is a maximal filter, i.e. there exists no filter F' such that $F \subsetneqq F'$.

Again it is easy to show that a principal filter over Σ^* is an ultrafilter iff it is of the form $F_{\{u\}}$ for some $u \in \Sigma^*$.

Definition 7. A filter X is a recognizable (well-recognizable) filter if the family X is recognizable (well-recognizable). Analogically we define a recognizable, resp. well-recognizable ultrafilter.

<u>Theorem 8</u>. A filter over Σ^* is recognizable iff it is a principal filter of the form F_L where L is a regular language.

<u>Proof</u>. First we show that every principal filter is self-compatible.

Let $L' \in C(F_L)$, for the sake of contradiction we shall assume that $L' \notin F_L$, i.e. there exists $u \in L$ such that $u \notin L'$. By the definition of C-closure there must exist $L_u \in F_L$ such that particularly $\Lambda \in Fst_\Lambda(\partial_u L') \equiv \Lambda \in Fst_\Lambda(\partial_u L_u)$ and thus $u \in L' \equiv u \in L_u$. But $u \in L_u$ because $L \subseteq L_u$ and thus also $u \in L'$, which contradicts the assumption.

Furthermore, for any $u \in \Sigma^*$,

$$\partial_u F_L = \partial_u \{ L' ; L \subseteq L' \} = \{ L'' ; \partial_u L \subseteq L'' \} .$$

Thus $\partial_u F_L = \partial_v F_L \equiv \partial_u L = \partial_v L$, i.e., F_L is a finitely derivable family iff L is a regular language.

Now we have known that a principal filter F_L is recognizable iff L is a regular language. It remains to show that every recognizable filter F must be principal, i.e. that $\bigcap F \in F$. Let F be a recognizable filter. First we show that if $\bigcap F \subseteq L$ and L is a complete language then $L \in F$ (for the definition of a complete language see e.g. [5], p. 47). In our notation L is complete language iff $(\forall u \in \Sigma^*)(\Sigma \subseteq Fst_\Lambda(\partial_u L))$. For $u \in L$,

$$Fst_\Lambda(\partial_u L) = \Sigma_\Lambda = Fst_\Lambda(\partial_u \Sigma^*)$$

and for $u \notin L$,

$$Fst_{\Lambda}(\partial_u L) = \Sigma = Fst_{\Lambda}(\partial_u(\Sigma^* - \{u\})).$$

But necessarily $\Sigma^* \in F$ (F is non-empty) and if $u \notin L$ then by the assumption $u \notin \bigcap F$, i.e. there exists $L' \in F$ such that $u \notin L'$ and since $L' \subseteq \Sigma^* - \{u\}$ then by the property 3) of filter also $\Sigma^* - \{u\} \in F$. Therefore $L \in C(F)$ and thus $L \in F$ by the assumption about recognizability of F. Now it is easy to choose arbitrary two complete languages L_1 and L_2 for which $L_1 \cap L_2 = \bigcap F$.

We have shown that $L_1 \in F$ and $L_2 \in F$ and thus also $L_1 \cap L_2 = \bigcap F \in F$ (property 2)).

<u>Theorem 9</u>. A principal filter of the form F_L is well-recognizable iff it is an ultrafilter.

<u>Proof</u>. We have stated (cf. [4],p. 196) that principal filter is an ultrafilter iff it is of the form $F_{\{u\}}$ for $u \in \Sigma^*$. By the preceding theorem $F_{\{u\}}$ is recognizable. Clearly for every $v \in \Sigma^*$ such that $lg(v) > lg(u)$, $\partial_v \overline{F_{\{u\}}} = \mathcal{L}(\Sigma)$. Thus $\overline{F_{\{u\}}}$ is finitely derivable and furthermore $C(\overline{F_{\{u\}}}) = \overline{F_{\{u\}}}$ because for every $L \in F_{\{u\}}$, $\Lambda \in Fst_{\Lambda}(\partial_u L)$, while for any $L' \in \overline{F_{\{u\}}}$, $\Lambda \notin Fst_{\Lambda}(\partial_u L')$. Thus also $\overline{F_{\{u\}}}$ is recognizable and so $F_{\{u\}}$ is a well-recognizable family.

Now let us assume, for contradiction, that F_L is not an ultrafilter, i.e. there exists $v, w \in L$ such that $w \neq v$. Thus by the definition of F_L we have $\Sigma^* - \{v\} \in \overline{F_L}$ and $\Sigma^* - \{w\} \in \overline{F_L}$. But for any $u \in \Sigma^*$ we have $u \neq v \Longrightarrow$

$$\Longrightarrow Fst_{\Lambda}(\partial_u \Sigma^*) = \Sigma_{\Lambda} = Fst_{\Lambda}(\partial_u(\Sigma^* - \{v\}));$$

$$u = v \Longrightarrow Fst_{\Lambda}(\partial_u \Sigma^*) = \Sigma_{\Lambda} = Fst_{\Lambda}(\partial_u(\Sigma^* - \{w\})).$$

Thus $\Sigma^* \in C(\overline{F_L})$ and since $\Sigma^* \notin \overline{F_L}$ we have $C(\overline{F_L}) \neq \overline{F_L}$ and so F_L is not a well-recognizable filter.

Q.e.d.

In the paper [2] we have shown that to every nontrivial well-recognizable family X there exists exactly one string $u_X \in \Sigma^*$ such that the families $\partial_v X$ are trivial (i.e. $\emptyset$ or $\mathcal{L}(\Sigma)$) for all $v \not\leq u_X$ while they are nontrivial and mutually distinct for all $v \leq u_X$. We have called u_X the characteristic string of a family X because it uniquely determines X regarding the algebraic decomposition of X to a finite number of basic families and regarding the (minimal) number of states of a branching automaton recognizing X. It can be easily seen that for an ultrafilter $F_{\{u\}}$, the string u satisfies the above conditions and thus $u_{F_{\{u\}}} = u$ (i.e. there exists finite branching automaton with $(lg(u) + 2)$ states recognizing the family $F_{\{u\}}$ - cf. [2]).

The preceding theorems showed us an interesting relationship between recognizable families and filters, as well as between well-recognizable families and ultrafilters.

We shall now turn to a dual notion to that of a filter, namely the ideal. We obtain results analogical to those concerning filters. Our definition of an ideal is a slight modification of that from [6], p. 132.

Definition 10. A non-empty set I of subsets of Σ^* is an ideal over Σ^* if

1) $\Sigma^* \notin I$;

2) if $A, B \in F$ then $A \cup B \in I$;

3) if $A \in I$ and $B \subseteq A$ then $B \in I$.

Again we shall call ideals over Σ^* simply ideals.

We want to talk about recognizable ideals. However, since always $\emptyset \in I$ no ideal is a "family" in our sense. We shall therefore use the following definition.

<u>Definition 11</u>. We say that an ideal I is a recognizable ideal if $I - \{\emptyset\}$ is a recognizable family of languages.

Similarly as in the case of principal filters we have again principal ideals of the form $I_A = \{B;\ B \subseteq A\}$, where $A \subsetneqq \Sigma^*$. An ideal is principal iff $\bigcup I \in I$.

<u>Theorem 12</u> . An ideal I is recognizable iff it is a principal ideal of the form I_A, where A is a regular language (possibly empty), $A \neq \Sigma^*$.

<u>Proof</u>. If $A = \emptyset$, $I_A - \{\emptyset\} = \emptyset$ is a trivial recognizable family. If $A = L \in \mathcal{L}(\Sigma)$, then in the same way as in Theorem 8 one can show that $I_L - \{\emptyset\}$ is self-compatible, as well as that it is finitely derivable iff L is finitely derivable.

It suffices to show that a recognizable ideal is principal, i.e. that $\bigcup I \in I$.

If $\bigcup I = \emptyset$ then $I = I_{\emptyset}$ is principal.

Otherwise we put $\bigcup I = L$ and show that L is in the C-closure of $I - \{\emptyset\}$. Since for every $L' \in I$, $L' \subseteq L$ and since an ideal is closed under finite union, for every $u \in \Sigma^*$ there surely exists $L_u \in I$ satisfying the conditions:

a) $(\forall v \in \Sigma^*)[\lg(v) = \lg(u) + 1 \Longrightarrow (v \in \mathrm{Pref}(L) \equiv v \in \in \mathrm{Pref}(L_u))]$;

b) $u \in L \equiv u \in L_u$.

However, then $Fst_{\Lambda}(\partial_u L) = Fst_{\Lambda}(\partial_u L_u)$. Thus $L \in C(I - \{\emptyset\}) = I - \{\emptyset\}$, i.e. I is a principal ideal.

Q.e.d.

R e f e r e n c e s

[1] HAVEL I.M.: Finite Branching Automata, Kybernetika 10 (1974), 281-302.

[2] BENDA V., BENDOVÁ K.: On Families Recognizable by Finite Branching Automata (in preparation).

[3] KATĚTOV M.: O základech matematického vyjadřování plánu, Mimeographed report, Faculty of Mathematics and Physics, Charles University, 1974.

[4] MALCEV A.I.: Algebraičeskije sistemy, Nauka, Moscow 1970.

[5] EILENBERG S.: Automata, Languages and Machines, Vol.A, Academic Press, New York 1974.

[6] VOPĚNKA P., HÁJEK P.: The Theory of Semisets, Academia, Prague and North-Holland, Amsterdam, 1972.

Výzkumný ústav železniční
U lužického semináře 3
11000 Praha 1
Československo

Matematický ústav ČSAV
Žitná 25
11567 Praha 1
Československo

(Oblatum 16.2. 1976)