INFORMATIQUE THÉORIQUE ET APPLICATIONS

TOM HEAD

The topological structure of adherences of regular languages

Informatique théorique et applications, tome 20, nº 1 (1986), p. 31-41
<http://www.numdam.org/item?id=ITA_1986__20_1_31_0>

© AFCET, 1986, tous droits réservés.

L'accès aux archives de la revue « Informatique théorique et applications » implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

Informatique théorique et Applications/Theoretical Informatics and Applications
(vol. 20, n° 1, 1986, p. 31 à 41)

THE TOPOLOGICAL STRUCTURE OF ADHERENCES OF REGULAR LANGUAGES (*) (**)

by Tom HEAD (¹)

Communicated by J.-E. PIN

Abstract. — *The topological structure of the adherence of a regular language is characterized by a finite invariant structure that can be algorithmically constructed from any automaton recognizing the language. Homeomorphism of the adherences of regular languages can be decided by constructing and comparing their associated invariant structures. These results are obtained by applying concepts and analyses given by R. S. Pierce in his study of zero-dimensional compact metric spaces of finite type.*

Résumé. — *La structure topologique de l'adhérence d'un langage rationnel est caractérisée par une structure invariante finie qui peut être construite par un algorithme à partir d'un automate reconnaissant le langage. L'homéomorphie de deux adhérences de langages rationnels peut être décidée en construisant et en comparant les structures invariantes qui leur sont associées. Ces résultats sont obtenus en employant des concepts et analyses développées par R. S. Pierce dans son étude des espaces métriques compacts de dimension zéro et de type fini.*

1. INTRODUCTION

Let A be a finite non-empty set. Let A^* be the set of finite *strings* of elements of A and A^ω the set of all infinite *sequences* of elements of A. With each language $L \subseteq A^*$, Boasson and Nivat [1] have associated a set Adh $L \subseteq A^\omega$ which they have called the *adherence* of L. A sequence s is in Adh L if every finite initial segment of s is an initial segment of some string in L. Equivalently, an s in A^ω is in Adh L if s is in the closure of L taken with respect to a natural metric topology for $A^* \cup A^\omega$ ([1], sec. IV). Since Adh L is a subset of this metric space, Adh L is itself a metric space. The purpose of this article is to give incisive descriptions of the topological structure of the spaces Adh L where L is a regular language.

In section 2 pertinent concepts of general topology are recalled and the definitions required for the metric space approach to adherences are given.

(*) Received in August 1984, revised in January 1985.

(**) This research was supported in part by Grant MCS-8303922 of the National Science Foundation of the United State of America.

(¹) Department of Mathematical Sciences, University of Alaska, Fairbanks, Alaska 99701, U.S.A.

ISSN en cours 86/01 31 11/$ 3.10/© Gauthier-Villars.

In Theorem 1 the topological spaces that arise as adherences of arbitrary languages are characterized abstractly as the zero-dimensional compact metrizable spaces and concretely as the closed subspaces of Cantor's ternary set. These elementary characterizations provide a general context for the detailed study of the adherences of regular languages in section 4.

R. S. Pierce has defined and studied zero-dimensional compact metric spaces of *finite type* in [5]. This is exactly the class of spaces that arise as adherences of regular languages. In section 3 an exposition of selected concepts from ([5], part I) is given. Some liberties have been taken in this exposition from the desire to make the specific applications given here quickly accessible. A finite invariant structure, here called a *structural diagram*, is associated with each space of the type under consideration. Pierce's result that the structural diagrams for these spaces characterize their topological structure is stated without proof. It provides a crutial decision procedure in section 4.

The critical results of this article are demonstrated in section 4 : For any regular language L, the topological space Adh L is of finite type and from any automaton recognizing L the structural diagram of Adh L is algorithmically constructible (theorem 2). Homeomorphism of adherences is decidable for regular languages (theorem 3).

This material was presented at the Ecole de Printemps d'Informatique Theorique, le Mont Dore, May 1984. A closely related version containing more expository material concerning the topological aspects of this subject appears in [2].

2. CHARACTERIZATIONS OF THE SPACES WHICH ARISE AS ADHERENCES

Let A be a finite non-empty set and let $A^\infty = A^* \cup A^\omega$. A metric d is defined on A^∞ as follows: Let u and v be in A^∞. If $u=v$ then $d(u, v)=0$. If $u \neq v$ and neither u nor v is an initial segment of the other then $d(u, v)=3^{-n}$ where n is the least positive integer for which $u(n) \neq v(n)$. If $u \neq v$ and one of u and v is an initial segment of the other then $d(u, v)=3^{-(n+1)}$ where n is the length of the shorter of u and v.

With respect to the metric d, both A^∞ and A^ω are compact ([1], [2]). For each language L in A^*, let L^- be the closure of L in A^∞. The *adherence* Adh L of L may be defined to be $A^\omega \cap L^-$. Since A^ω is closed in A^∞, Adh L is a closed subset of both A^ω and A^∞. As closed subspaces of the compact space A^∞, both L^- and Adh L are compact.

Recall that a topological space is *zero-dimensional* if the space has a basis consisting of sets each of which is closed as well as open. The collection of spheres :

$$S_{\infty}(u, 3^{-n})=\{v \text{ in } A^{\infty}: d(u, v)<3^{-n}\}=\{v \text{ in } A^{\infty}: d(u, v)\leqq 3^{-(n+1)}\},$$

where u is in A^{∞} and n is a positive integer, is a basis for the topology of A^{∞}. Since for each u in A^{∞}, $d(u, \_)$ is continuous it follows that each sphere $S_{\infty}(u, 3^{-n})$ is both open and closed. Consequently A^{∞} and all of its subspaces are zero-dimensional.

Recall that a topological space is *perfect* if no open subset is a singleton, i. e. if the space has no isolated points. Suppose now that A contains at least two elements. Then each non-empty open subset of A^{ω} contains a sphere $S_{\omega}(u, 3^{-n})=S_{\infty}(u, 3^{-n})\cap A^{\omega}=\{v \text{ in } A^{\omega}: d(u, v)<3^{-n}\}$, for some u in A^{ω} and some positive integer n, and each such sphere has cardinal $2^{\aleph_0}$. Consequently A^{ω} is perfect. For each u in A^*, $\{u\}$ is an open subset of A^{∞} and consequently A^{∞} is not perfect.

The classical example of a perfect zero-dimensional compact space is the ternary set introduced by G. Cantor. It is the subspace of the closed interval [0, 1] consisting of those real numbers r possessing a representation in the base 3 system of numeration for which only the digits 0 and 2 appear. We will call this space *the Cantor space* C. The following theorem is a classical topological characterization of C : *Every non-empty perfect zero-dimensional compact metric space is homeomorphic with the Cantor space.* The following corollary provides a topological characterization of the closed subspaces of C: *Every zero-dimensional compact metric space is homeomorphic with a subspace of the Cantor space.* See Hocking and Young ([3], p. 100) for proofs of these theorems and Willard ([6], p. 315) for historical references.

We have observed that for any finite set A with at least two elements, A^{ω} is a perfect zero-dimensional compact metric space. Consequently A^{ω} *is homeomorphic with* C. Since for any language L in A^*, we have observed that Adh L is zero-dimensional and compact, to conclude that Adh L is homeomorphic with C it is necessary and sufficient to show that Adh L is perfect. In this way it can be verified that the Dyck languages and several other common examples of context-free languages have adherences homeomorphic with C. On the other hand Adh $a^* b^*$ contains b^{ω} and $\{b^{\omega}\}$ is an open set. Thus Adh $a^* b^*$ is not homeomorphic with C.

THEOREM 1. — *Let S be a topological space. Then the following conditions are equivalent:*

(1) *There is a language L over an alphabet A such that S is homeomorphic with* Adh L.

(2) *S is homeomorphic with a closed subset of the Cantor space C.*

(3) *S is zero-dimensional, compact, and metrizable.*

Proof. $(1 \to 2)$: Let A be a finite non-empty set. After adjoining an additional symbol to A if necessary, we may assume that A contains at least two symbols. Then A^ω is homeomorphic with C. Thus if S is homeomorphic with Adh L for some $L \subseteq A^*$, then S is homeomorphic with a closed subset of C.

$(2 \to 1)$: Since C and $\{0, 2\}^\omega$ are homeomorphic we may suppose that S is a closed subset of $\{0, 2\}^\omega$. For $L = \{u$ in $\{0, 2\}^*$: there is a v in $\{0,2\}^\omega$ for which uv is in $S\}$ we have $S = \text{Adh } L$.

$(2 \to 3)$: C and all its closed subsets are zero-dimensional, compact, and metrizable.

$(3 \to 2)$: This follows from the second of the two results stated in the fifth paragraph of this section. □

3. SPACES OF FINITE TYPE AND THEIR STRUCTURAL DIAGRAMS

Let X be a topological space. Let $\mathbf{P}(X)$ be the set of all subsets of X. Regard $\mathbf{P}(X)$ as an algebra with seven operations: $\varnothing$, X, $\cup$, $\cap$, ^{C}, ${}^-$, $'$. The first five operations are the usual Boolean operations: $\varnothing$ and X are the nullary operations of specifying the empty set and the universal set X for $\mathbf{P}(X)$, $\cup$ and $\cap$ are the binary operations of union and intersection, and ^{C} is the unary operation of complementation. The final two operations are unary. For each Y in $\mathbf{P}(X)$:

Y^- is the topological closure of Y; and

Y' is the topological derivative of Y which may be defined by $Y' = \{y$ in $Y : y$ is in $(Y \setminus \{y\})^-\}$. Equivalently, Y' is the set of all points in Y that are not isolated.

By a subalgebra of $\mathbf{P}(X)$ we mean a family of subsets of X that is closed under the seven operations of $\mathbf{P}(X)$. From the nature of nullary operations every subalgebra must contain $\{\varnothing, X\}$. Since the intersection of every collection of subalgebras is a subalgebra, $\mathbf{P}(X)$ contains a unique minimal subalgebra that necessarily contains $\{\varnothing, X\}$. *Let* $\mathbf{B}(X)$ *be the minimal subalgebra of* $\mathbf{P}(X)$.

DEFINITION : A topological space X is of *finite type* if $\mathbf{B}(X)$ *is finite.*

Let X be a topological space of finite type. Then $\mathbf{B}(X)$ *is a finite Boolean algebra with the two additional unary operations* ${}^-$ and $'$. As such, $\mathbf{B}(X)$ *has atoms, i. e. minimal non-empty subsets, and consists of* 2^n subsets where n is the number of its atoms. Let $\mathbf{A}(X)$ be the set of atoms of $\mathbf{B}(X)$.

LEMMA 1 : *Let A and B in* $\mathbf{A}(X)$ *and suppose that* $A \subseteq B^-$ *and* $B \subseteq A^-$. *Then* $A = B$.

Proof: Let $\mathbf{S} = \{Y$ in $\mathbf{B}(X)$: *either* Y contains both A and B or Y contains neither$\}$. Apparently $\varnothing$ and X are in $\mathbf{S}$. It is likewise elementary to verify that $\mathbf{S}$ is closed under $\cup$, $\cap$, ^{c}, ${}^-$, and $'$. Consequently $\mathbf{S}$ is a subalgebra of $\mathbf{B}(X)$ *and, by the minimality of* $\mathbf{B}(X)$, $\mathbf{S} = \mathbf{B}(X)$. *Since* A (resp. B) is in $\mathbf{B}(X)$ *and* $A \subseteq A$ (resp. $B \subseteq B$) it follows that $B \subseteq A$ (resp. $A \subseteq B$). □

We provide $\mathbf{A}(X)$ with a *partial order* $\leqq$ by defining, for A and B in $\mathbf{A}(X)$, $A \leqq B$ if $A \subseteq B^-$. The required anti-symmetry of $\leqq$ is provided by lemma 1. With each A in $\mathbf{A}(X)$ we associate the cardinal number, Card A, of the subset A of X.

DEFINITION: Let X be a topological space of finite type. The *structural diagram* of X is the partially ordered set $\mathbf{A}(X)$ together with the function Card: $\mathbf{A}(X) \rightarrow$ the set of those cardinal numbers which do not exceed the cardinal of X.

Spaces X and Y as in this definition have *isomorphic structural diagrams* if there is a bijection $i: \mathbf{A}(X) \rightarrow \mathbf{A}(Y)$ for which, for each A in $\mathbf{A}(X)$, Card $A = \text{Card } i(A)$ and, for each A and B in $\mathbf{A}(X)$, $A \leqq B$ if and only if $i(A) \leqq i(B)$.

In accordance with this isomorphism concept we consider a structural diagram to be merely a Hasse diagram with cardinal numbers attached to its nodes, i. e. we do not attend to the fact that each node represents a subspace of the original space. Thus comparing structural diagrams is a purely finitary activity.

The following theorem assures us of the importance of the structural diagram concept. It is also the key to deciding homeomorphism of adherences of regular languages as is shown in section 4. For a proof of this theorem *see* [5]. A major aspect of the proof is treated in [2].

The Structure Theorem of R. S. Pierce. Let X and Y be zero-dimensional compact metric spaces of finite type. Then X and Y are homeomorphic if and only if they have isomorphic structural diagrams.

4. THE TOPOLOGICAL STRUCTURE OF ADHERENCES OF REGULAR LANGUAGES

Let A be a finite non-empty set and let $L \subseteq A^*$ be a regular language. Let $G = (S, A, E, s_0, F)$ be a deterministic automaton that recognizes L, where S is the finite set of states, $E \subseteq S \times A \times S$ is the set of edges, s_0 is the initial state, and $F \subseteq S$ is the set of final states. Without loss of the generality of L we assume that s_0 cannot be entered, i. e. that there is no edge of the form

(s, a, s_0) in E. We dispose of the states and edges of G that are useless for the investigation of the structure of Adh L: Delete from S any state that is not accessible from s_0. Delete any state $s \neq s_0$ from S for which the language accepted by the automaton (S, A, E, s, F) is finite. Finally, delete from E all edges in which deleted states occur. Let S and E now denote the sets of states and edges *after* these deletions have been made. Let $H=(S, A, E, s_0, S)$ and let $L(H)$ be the language recognized by H. Observe first that, due to the way deletions of states and edges were made, $L(H)$ consists of the initial segments of the sequences in Adh L. ($L(H)$ is the *center* of L as defined in ([1], Def. 3)). Observe second that the infinite sequences in Adh L are precisely the sequences which are labels of infinite paths in H that start from s_0.

We have not yet reduced the set S of states sufficiently to allow the clearest exposition of our results. Notice that no difficulty is introduced if instead of demanding $E \subseteqq S \times A \times S$ in the definition of a finite automaton we demand only that E be a *finite* subset of $S \times A^+ \times S$. From this point on we employ this slightly more liberal concept of a finite automaton.

Relative to any linear ordering of $S \setminus \{s_0\}$ carry out the following deletions and insertions for each $s \neq s_0$: If there is no proper loop at s, i. e. if there is no string u in A^+ that labels a path that begins at s and ends at s, delete s from S. For each s that is so deleted, carry out the following insertions and deletions from E: If $(p_1, u_1, s), \ldots, (p_m, u_m, s)$ are all the edges ending at s and $(s, v_1, q_1), \ldots, (s, v_n, q_n)$ are all the edges beginning at s then add to E the edges $(p_i, u_i v_j, q_j)$ for $1 \leqq i \leqq m$ and $1 \leqq j \leqq n$. After these additions are made, delete all edges from E in which s occurs. Now let S and E denote the sets of states and edges *after* these deletions and insertions have been made. With these new settings of S and E, let $H=(S, A, E, s_0, S)$. The new $L(H)$ may be smaller than the previous version due to the deletion of this last set of states. However, the language of initial segments of this new $L(H)$ is identical with the previous version (i. e. the center is unchanged), and Adh L may still be described as the set of infinite sequences in A which are labels of paths in H that start from s_0.

The automaton $H=(S, A, E, s_0, S)$ is now the tool we need for the computation of the structural diagram of the topological space Adh L. We begin by imitating the construction of the algebra $\mathbf{P}(X)$ in section 3 with X replaced by $T=S \setminus \{s_0\}$. Let $\mathbf{P}(T)$ be the set of all subsets of T. Regard $\mathbf{P}(T)$ as an algebra with the seven operations: $\varnothing$, T, $\cup$, $\cap$, ^{C}, ^{K}, and ^{D}. The first five operations are the usual Boolean operations. The final two are unary. For each Y in $\mathbf{P}(T)$ we define:

$Y^K = \{s$ in T: Y is accessible from s, i. e. there is a path in H beginning at s and ending at a state in $Y\}$.

$Y^D = \{s$ in Y: there are at least two infinite paths beginning at s and eventually passing only through states in $Y\}$.

Notice that since there is a loop at every state in T, it follows that if s is in Y^K then there is an infinite sequence v that labels a path that begins at s and is frequently in Y.

The seven operation algebra $\mathbf{P}(T)$ must contain a unique minimal subalgebra that necessarily contains $\{\varnothing, T\}$. *Let* $\mathbf{D}(T)$ *be the minimal subalgebra of* $\mathbf{P}(T)$. Then $\mathbf{D}(T)$ is a finite Boolean algebra with two additional unary operations ^{K} and ^{D}. As such it has atoms and consists of 2^n subsets where n is the number of its atoms. Let $\mathbf{C}(T)$ be the set of atoms of $\mathbf{D}(T)$. An exact parallel of the proof of lemma 1 establishes the following result.

LEMMA 2: *Let A and B be in $\mathbf{C}(T)$ and suppose that $A \subseteq B^K$ and $B \subseteq A^K$. Then $A = B$.* □

We provide $\mathbf{C}(T)$ with a *partial order* $\leqq$ by defining, for A and B in $\mathbf{C}(T)$, $A \leqq B$ if $A \subseteq B^K$. The required anti-symmetry of $\leqq$ is provided by lemma 2.

For each s in T, let $\mathrm{In}(s) = \{u$ in Adh L: the path in H beginning at s_0 and labeled by u passes frequently, i. e. infinitely often, through $s\}$. For each subset X of T, let $\mathrm{In}(X) = \bigcup \{\mathrm{In}(s) : s$ in $X\}$. We have defined a function $\mathrm{In} : \mathbf{P}(T) \to \mathbf{P}(\mathrm{Adh}\, L)$. Although In is not a morphism of $\mathbf{P}(T)$ into $\mathbf{P}(\mathrm{Adh}\, L)$, we will show that when we restrict the domain of In to the subalgebra $\mathbf{D}(T)$ we obtain a morphism $\mathrm{In} : \mathbf{D}(T) \to \mathbf{P}\,\mathrm{Adh}\, L$ that maps $\mathbf{D}(T)$ isomorphically onto $\mathbf{B}(\mathrm{Adh}\, L)$. The following concept and lemma will be convenient for demonstrating this isomorphism.

A subset X of the set T is *coherent* if, for any pair p, q of mutually accessible states of H, either both p and q lie in X or neither lies in X. The family of coherent subsets of T is easily verified to be a subalgebra of $\mathbf{P}(T)$ and consequently by the minimality of $\mathbf{D}(T)$ we have:

LEMMA 3: *Each set in $\mathbf{D}(T)$ is coherent.* □

This lemma has the following helpful consequence: for any X in $\mathbf{D}(T)$, if an infinite path is frequently in X then it is eventually in X.

PROPOSITION: In: $\mathbf{D}(T) \to \mathbf{P}(\mathrm{Adh}\, L)$ *is a morphism of seven operation algebras that maps* $\mathbf{D}(T)$ *isomorphically onto* $\mathbf{B}(\mathrm{Adh}\, L)$.

Proof: Let X and Y be in $\mathbf{D}(T)$. Apparently $\mathrm{In}(\varnothing) = \varnothing$, $\mathrm{In}(\mathrm{T}) = \mathrm{Adh}\, L$, and $\mathrm{In}(X \cup Y) = \mathrm{In}(X) \cup \mathrm{In}(Y)$. That $\mathrm{In}(X \cap Y) \subseteq \mathrm{In}(X) \cap \mathrm{In}(Y)$ is also immediately clear. Suppose u is in $\mathrm{In}(X) \cap \mathrm{In}(Y)$. Such a u is *frequently* in both X and Y and therefore *eventualy* in both X and Y. Then u is eventually in

$X \cap Y$. Consequently $\mathrm{In}(X) \cap \mathrm{In}(Y) \subseteq \mathrm{In}(X \cap Y)$ and the the desired equality follows. Again by the coherence of X, every u in Adh L must lie in precisely one of $\mathrm{In}(X)$ and $\mathrm{In}(X^C)$ and consequently $\mathrm{In}(X^C) = \mathrm{In}(X)^C$.

Each of the four inclusions required to verify $\mathrm{In}(X^K) = \mathrm{In}(X)^-$ and $\mathrm{In}(X^D) = \mathrm{In}(X)'$ is given in a separate paragraph:

$\mathrm{In}(X^K) \subseteq \mathrm{In}(X)^-$: Let u be in $\mathrm{In}(X^K)$. Then there exists a state p in X^K for which: (1) $u = u_0 u_1 u_2 \ldots u_i \ldots$ where u_0 labels a path from s_0 to p and, for $i \geqq 1$, u_i labels a non-null loop from p to p; and (2) there is an infinite path beginning at p, labelled by a sequence v, which is eventually in X. Consequently for any open subset of Adh L that contains u there is a sufficiently large i so that $u_0 u_1 u_2 \ldots u_i v$ lies in the open set. Since, for all $j \geqq 0$, $u_0 u_1 u_2 \ldots u_j v$ lies in $\mathrm{In}(X)$, u is in $\mathrm{In}(X)^-$.

$\mathrm{In}(X)^- \subseteq \mathrm{In}(X^K)$: Let u be in $\mathrm{In}(X)^-$. Let p be a state for which u is in $\mathrm{In}(p)$. Then $u = u_0 v$ where u_0 labels a path from s_0 to p and v labels an infinite path that is frequently at p. Since u is in $\mathrm{In}(X)^-$ there is a sequence w for which $u_0 w$ is in $\mathrm{In}(X)$. Then w labels a path that begins at p and is eventually in X. Consequently, p is in X^K and u is in $\mathrm{In}(X^K)$.

$\mathrm{In}(X^D) \subseteq \mathrm{In}(X)'$: Let u be in $\mathrm{In}(X^D)$. Then there exists a state p in $X^D \subseteq X$ for which: (1) $u = u_0 u_1 u_2 \ldots u_i \ldots$ where u_0 labels a path from s_0 to p, and for $i \geqq 1$, u_i labels a non-null loop from p to p; and (2) there are infinite paths beginning at p, labeled by distinct sequences v and w, which lie entirely in X. Consequently for any open subset of Adh L that contains u there is a sufficiently large i so that $u_0 u_1 u_2 \ldots u_i v$ and $u_0 u_1 u_2 \ldots u_i w$ lie in the open set. Since at least one of these sequences is different from u, the open set cannot be a singleton. Thus u is not isolated in $\mathrm{In}(X)$, i. e. u is in $\mathrm{In}(X)'$.

$\mathrm{In}(X)' \subseteq \mathrm{In}(X^D)$: Let u be in $\mathrm{In}(X)'$. Then there exists a state p in X for which: (1) $u = u_0 v$ where u_0 labels a path from s_0 to p and v labels an infinite path lying entirely in X; and (2) there is a sequence $u_0 w$ in $\mathrm{In}(X)$ for which $u_0 w \neq u$. Then v and w label distinct paths beginning at p and passing only through states in X. Consequently p is in X^D and u is is $\mathrm{In}(X^D)$.

The verification that In is a morphism is now complete. The image of In is a subalgebra of $\mathbf{P}(\mathrm{Adh}\ L)$ and therefore must contain the minimal subalgebra of $\mathbf{P}(\mathrm{Adh}\ L)$ which is $\mathbf{B}(\mathrm{Adh}\ L)$. The complete inverse image of $\mathbf{B}(\mathrm{Adh}\ L)$ with respect to In is a subalgebra of $\mathbf{D}(T)$ and therefore must be $\mathbf{D}(T)$ itself by the minimality of $\mathbf{D}(T)$. Thus the image of In is $\mathbf{B}(\mathrm{Adh}\ L)$.

From the constructions that lead to the specific form of $H = (S, A, E, s_0, S)$ in use here, for every s in $T = S \setminus \{s_0\}$, s is accessible from s_0 by a path with some label u in A^+, and there is a proper loop at s labeled with some label v in A^+. Consequently uv^ω is in $\mathrm{In}(s)$ which is therefore not empty. It follows that the only subset X of T for which $\text{In}(X)=\varnothing$ is $X=\varnothing$. Since In is a homomorphism of Boolean algebras, it follows that In is an injective function. □

THEOREM 2: *The adherence of a regular language is a zero-dimensional compact metric space of finite type and its structural diagram is algorithmically constructible.*

Proof: Let L be a regular language. From section 2 Adh L is a zero-dimensional compact metric space. From a deterministic automaton recognizing L construct the automaton $H=(S, A, E, s_0, S)$ as specified in this section. By proposition 1 the function In provides an isomorphism of $\mathbf{D}(T)$ with $\mathbf{B}(\text{Adh } L)$ where $T=S\setminus\{s_0\}$. Since $\mathbf{D}(T)$ is finite, $\mathbf{B}(\text{Adh } L)$ is finite and Adh L is of finite type.

The construction of the structural diagram of Adh L is outlined as follows: Construct $\mathbf{D}(T)$. Since In maps $\mathbf{D}(T)$ isomorphically onto $\mathbf{B}(\text{Adh } L)$, it follows that In also maps the partially ordered set $\mathbf{C}(T)$ isomorphically onto the partially ordered set $\mathbf{A}(\text{Adh } L)$. From $\mathbf{D}(T)$ construct $\mathbf{C}(T)$. Apply In to $\mathbf{C}(T)$ to produce the partially ordered set $\mathbf{A}(\text{Adh } L)$. In the remaining paragraph we consider the details:

For p in T let L_p and L_{pp} be the languages recognized by $(S, A, E, s_0, \{p\})$ and $(S, A, E, p, \{p\})$, respectively. The languages L_p and L_{pp} may be expressed as regular expressions and consequently $\text{In}(p)=L_p L_{pp}^{\omega}$ is expressible in terms of regular expressions. Thus, for each X in $\mathbf{D}(T)$, $\text{In}(X)=\bigcup\{L_p L_{pp}^{\omega}; p \text{ in } X\}$ where all the L_p and L_{pp} are regular expressions. From such a representation of $\text{In}(X)$, Card $\text{In}(X)$ is easily read off. (We remark that it is also not difficult to compute the required cardinals directly from $\mathbf{C}(T)$, consulting H only to determine the exact value of the finite cardinals). □

THEOREM 3: *Homeomorphism of adherences is decidable for regular languages.*

Proof: Let M and N be constructively given regular languages. Construct the structural diagrams of Adh M and Adh N as in theorem 2. Determine whether these structural diagrams are isomorphic. By the Structure Theorem of R. S. Pierce, Adh M and Adh N are homeomorphic precisely if these structural diagrams are isomorphic. □

THEOREM 4: *Every zero-dimensional compact metrizable space of finite type is homeomorphic with the adherence of a regular language. The language may be chosen to be two-testable in the strict sence.*

DISCUSSION: Let the structural diagram of such a space be given. Construct the underlying multi-graph of an automaton as follows: Direct the edges of the Hasse diagram underlying the structural diagram from smaller to larger. To each node of the diagram attach a simple directed loop involving no other node. At those nodes for which the associated cardinal number in the structural diagram is $2^{\aleph_0}$, attach a second simple directed loop. Adjoin a new node s_0 (to be used as a start state). Treat the nodes for which the associated cardinal is finite as follows: Let p be a node for which the associated cardinal is the positive integer n. Insert n distinct edges from s_0 to the node p. Treat any additional minimal node q as follows: Insert one edge from s_0 to q.

We now have the underlying directed multi-graph of the automaton we require. There is great freedom in choosing an alphabet and in assigning symbols of the alphabet to edges. At one extreme one can use a different symbol for every edge of the directed graph, thus using an alphabet equal in size to the number of edges. This choice insures that the language generated will be two-testable in the strict sense as defined in [4], p. 17. At the other extreme one can use a two symbol alphabet, say $A=\{a, b\}$, and associate with the edges not single symbols but strings from A^k where k is chosen as required by the largest outdegree occurring at the nodes of the directed graph. Let all states be final states. Whichever choice is made, the adherence of the language recognized by such an automaton is homeomorphic with the original space since our construction procedure has guaranteed that they have isomorphic structural diagrams. See [2] for further details. □

ACKNOWLEDGMENT

The author wishes to express his thanks to his colleague Robert Piacenza for valuable conversations concerning topological questions and particularly for pointing out the Memoir by R. S. Pierce [5] which was so fundamental to this present work.

REFERENCES

1. L. BOASSON and M. NIVAT, *Adherences of Languages*, Journal of Computer and System Sciences, Vol. 20, 1980, pp. 285-309.
2. T. HEAD, *The Adherences of Languages as Topological Spaces*, in: *Automata on Infinite Wonds*, edited by M. Nivat and D. Perrin, LNCS, vol. 192, Springer-Verlag, 1985.
3. J. G. HOCKING and G. S. YOUNG, *Topology*, Addison-Wesley, Reading, Mass., U.S.A., 1961.

4. R. McNaughton and S. Papert, *Counter-Free Automata*, M.I.T. Press, Cambridge, Mass., U.S.A., 1971.
5. R. S. Pierce, *Compact zero-dimensional metric spaces of finite type*, Memoirs of the Amercan Mathematical Society, No. 130, Providence, Rhode Island, U.S.A., 1972.
6. S. Willard, *General Topology*, Addison-Wesley, Reading, Mass., U.S.A., 1970.

Note added in proof: The following are recommended further references.

R. Lindner and L. Staiger, *Algebraische Codierungstheorie - Theorie der sequentiellen Codierungen*, Akademie-Verlag, Berlin, 1977.

L. Staiger, *Finite-state ω-languages*, J. Comput. System Sci. 27 (1983), pp. 434-448.