SOME APPLICATIONS OF LOGICAL SYNTAX TO DIGITAL COMPUTER PROGRAMMING

A thesis presented

by

Richard Manning Karp

to

The Division of Engineering and Applied Physics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

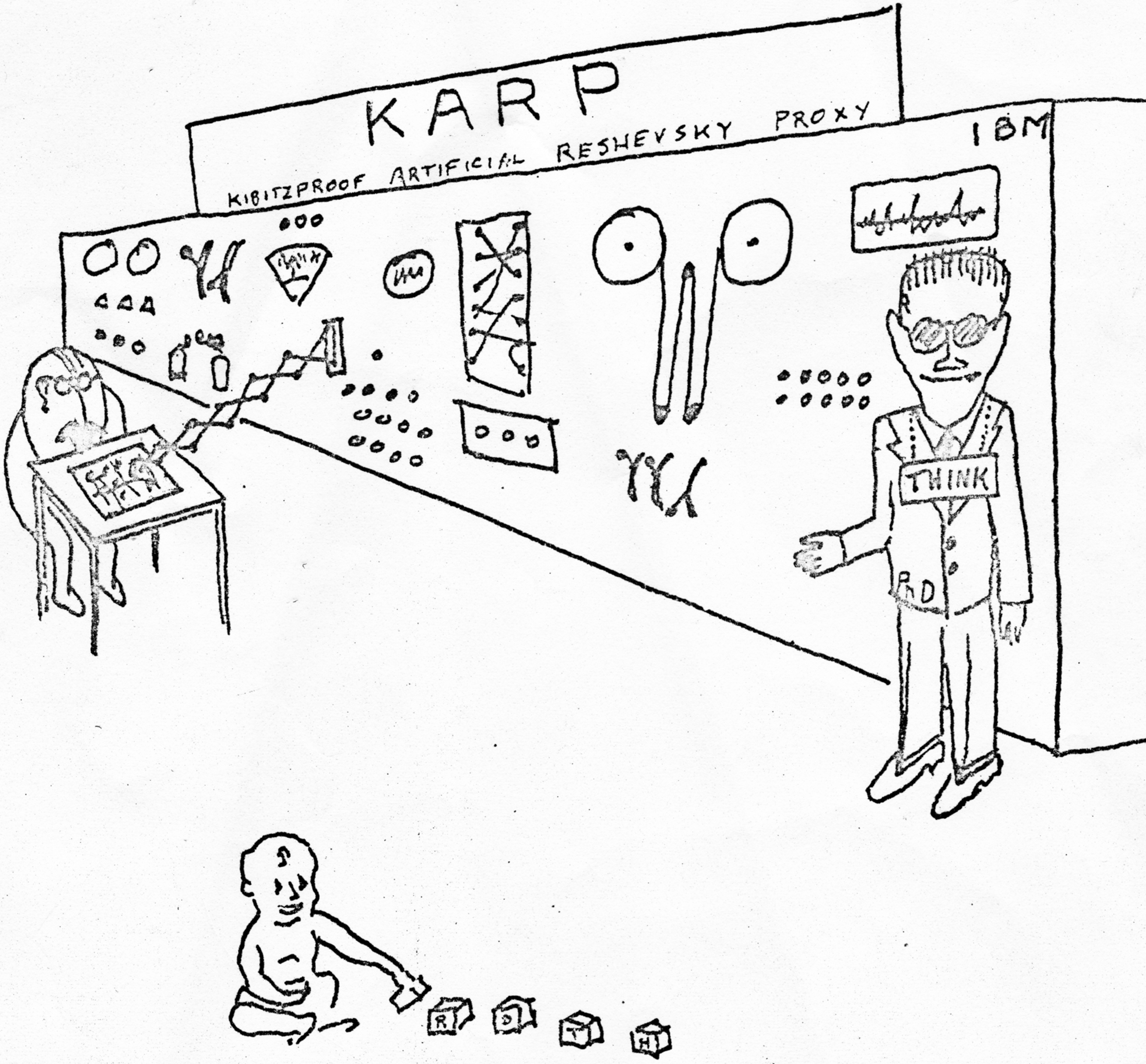
Applied Mathematics

Harvard University

Cambridge, Massachusetts

January 1959

FRONTISPIECE



000	11 48031 480	030	C 057------
001	31 54031 600	031	B 056C 055
002	30 660B 053	032	B 057C 056
003	C 054K	033	------U 019
004	[B 480T1 007]	034	------------
005	[S- 480J 055]	035	------------
006	------U 008	036	[B 056C 720]
007	C 055------	037	B 054L 058
008	K C 002	038	------Q8 040
009	[B 600T2 012]	039	------U 043
010	[S- 600J 056]	040	52 72052 780
011	------U 013	041	61 62
012	------C 056	042	------999999
013	B 056L 055	043	B 004A- 059
014	------T3 016	044	C 004B 005
015	------U 019	045	A- 059C 005
016	B 055C 057	046	B 009A- 059
017	B 056C 055	047	C 009B 010
018	B 057C 056	048	A- 059C 010
019	L 056B 053	049	B 036A- 053
020	------Q4 036	050	C 036B 054
021	D- 053J 001	051	A- 053C 054
022	L 001P- 055	052	------U 004
023	06 C 057	053	-----------1
024	B 05605	054	------------
025	K M- 057	055	------------
026	S- 055;1	056	------------
027	.1 L 056	057	------------
028	------Q5 036	058	000000000100
029	L 002Q6 036	059	000001000000

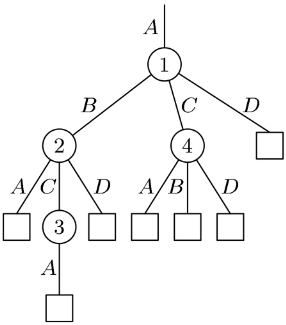
A Program for Executing the Euclidean Algorithm

Fig. 3-4

29. [*M47*] Exercise 28 shows that the polyphase distribution is optimal among all merge-until-empty patterns in the minimum-phase sense. But is it optimal also in the minimum-pass sense?

Let a be relatively prime to b, and assume that $a+b$ is the Fibonacci number F_n. Prove or disprove the following conjecture due to R. M. Karp: The number of initial runs processed during the merge-until-empty pattern starting with distribution (a, b) is greater than or equal to $((n-5)F_{n+1} + (2n+2)F_n)/5$. (The latter figure is achieved when $a = F_{n-1}$, $b = F_{n-2}$.)

The determination of strictly optimum T-tape merge patterns — that is, of T-lifo trees whose path length is minimum for a given number of external nodes — seems to be quite difficult. For example, the following nonobvious pattern turns out to be an optimum way to merge seven initial runs on four tapes, reading backwards:



(8)

A one-way merge is actually necessary to achieve the optimum! (See exercise 8.) On the other hand, it is not so difficult to give constructions that are *asymptotically* optimal, for any fixed T.

Let $K_T(n)$ be the minimum external path length achievable in a T-lifo tree with n external nodes. From the theory developed in Section 2.3.4.5, it is not difficult to prove that

$$K_T(n) \geq nq - \lfloor ((T-1)^q - n)/(T-2) \rfloor, \qquad q = \lceil \log_{T-1} n \rceil, \tag{9}$$

since this is the minimum external path length of *any* tree with n external nodes and all nodes of degree $< T$. At the present time comparatively few values of $K_T(n)$ are known exactly. Here are some upper bounds that are probably exact:

$n =$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$K_3(n) \leq$	0	2	5	9	12	16	21	25	30	34	39	45	50	56	61
$K_4(n) \leq$	0	2	3	6	8	11	14	17	20	24	27	31	33	37	40

(10)

Karp discovered that *any* tree whose internal nodes have degrees $< T$ is *almost* T-lifo, in the sense that it can be made T-lifo by changing some of the external nodes to one-way merges. In fact, the construction of a suitable labeling is fairly simple. Let A be a particular tape name, and proceed as follows:

Dec. 6, 1971

Norman Zadeh
Bostwick Wyman
Peter Brucker
M. A. Pollatschek
Harold Brown
[illegible]
[illegible]
Richard Karp
Daniel Gale
Donald Knuth
Russell N Taylor
[illegible]
John Mitchem
Henry D. Friedman
David A. Klarner
Harold S. Stone
Richard Stanley
Ashok Chandra
Bob Tarjan
Dick Sweet
LARRY MASINTER
Larry Carter
John T. Gill III
Ronald Fagin
Zohar Manna
Michael Fredman
Hal Gabow
Peter Deutsch
Barbara Goss
David Luckham
RW Floyd
R. Weyhrauch
M Newey
B. Elspas
Ron Rivest
Elwyn Berlekamp
Forest Baskett
L. Guibas

Dec 6 1971

MW Green (SRI)
Robin Milner
Vaughan Pratt

R. M. Karp

'Reducibilities between combinatorial programs'

Combinatorial Seminar – 9th February 1976

'Probabalistic Analysis of a Heuristic Set Covering Algorithm'

by R. M. Karp

John R. Gilbert
David V. Wall
Howard Johnston
Gerald [illegible]
Ed J. Gilbert
Scott E Kim
Sam Bent
Dereck Zave
Vic Klee
Don Woods
[illegible]
Dick Karp
John C Beatty
[illegible]
James Hunt
Janet R Roberts
Leo Guibas
Terry Roberts
Sandra Forsythe
Kelly Booth (Chatman)
Ramez El-Masri
John Reiser
Lyle Ramshaw
Tom Schaefer
Thomas Lengauer
G. Polya
scot [illegible]
D. Avis
J. Zolnowsky
Mark R Brown
M. Paterson
V. [illegible]
Bengt Aspvall

The Transitive Closure of a Random Digraph

Richard M. Karp*
University of California at Berkeley, Berkeley, CA 94720 and International Computer Science Institute, Berkeley, CA 94720

ABSTRACT

In a random n-vertex digraph, each arc is present with probability p, independently of the presence or absence of other arcs. We investigate the structure of the strong components of a random digraph and present an algorithm for the construction of the transitive closure of a random digraph. We show that, when n is large and np is equal to a constant c greater than 1, it is very likely that all but one of the strong components are very small, and that the unique large strong component contains about $\Theta^2 n$ vertices, where Θ is the unique root in $[0, 1]$ of the equation $1 - x - e^{-cx} = 0$. Nearly all the vertices outside the large strong component line in strong components of size 1. Provided that the expected degree of a vertex is bounded away from 1, our transitive closure algorithm runs in expected time $O(n)$. For all choices of n and p, the expected execution time of the algorithm is $O(w(n)(n \log n)^{4/3})$, where $w(n)$ is an arbitrary nondecreasing unbounded function. To circumvent the fact that the size of the transitive closure may be $\Omega(n^2)$ the algorithm presents the transitive closure in the compact form $(A \times B) \cup C$, where A and B are sets of vertices, and C is a set of arcs.

1. INTRODUCTION

The probability space of digraphs $D_{n,p}$ is defined as follows: each point in the space is a digraph with vertex set $\{1, 2, \ldots, n\}$ having no loops or multiple arcs, and the probability of a given digraph D with e arcs is $p^e(1-p)^{n(n-1)-e}$. In other words, each arc is present with probability p, independently of the presence or absence of other arcs. We shall study the structure of the strongly connected

* Research supported by NSF Grants DCR-8411954 and CCR-8612563, and by the International Computer Science Institute, Berkeley, California.

Random Structures and Algorithms, Vol. 1, No. 1 (1990)

unanswered seems to be almost endless. But we shall close this list of research problems by stating what seems to be the single most important related area ripe for investigation at the present time. Wright [42] gave a procedure for computing the number of *strongly connected labeled digraphs* of excess r, analogous to his formulas for connected labeled undirected graphs. Random directed multigraphs are of great importance in computer applications, and it is shocking that so little attention has been given to their study so far. Karp [21] carried Wright's investigations further and discovered a beautiful theorem: A random digraph with $n(1+\mu)$ directed arcs almost surely has a giant strong component of size $\sim\Theta(\mu)^2 n$, when $\Theta(\mu)$ is the factor such that an undirected graph with $\frac{1}{2}n(1+\mu)$ edges almost surely has a giant component of size $\sim\Theta(\mu)n$. (The function $\Theta(\mu)$ is $(\mu+\sigma)/(1+\mu)$, according to (23.11). Karp's investigation was based on $\mathbf{D}_{n,p}$, in which every directed arc is present with probability p, but a similar result surely holds for other models of random digraphs.) A complete analysis of the random *directed* multigraph process is clearly called for, preferably based on generating functions so that extensive quantitative information can be derived without difficulty.

Here is a sketch of how such an investigation might begin. The *directed multigraph process* consists of adding directed arcs $x \rightarrow y$ repeatedly to an initially empty multiset of arcs on the vertices $\{1, 2, \ldots, n\}$, where x and y are independently and uniformly distributed between 1 and n. The *compensation factor* $\kappa(M)$ of a multidigraph M with m_{xy} arcs from x to y is $1/\Pi_{x=1}^{n}\Pi_{y=1}^{n} m_{xy}!$; we can use it to compute bivariate generating functions as in (2.1). The bgf for all possible multidigraphs is $\Sigma_{n\geq 0}\, e^{n^2 w} z^n/n! = G(2w, z)$.

Let $\mathscr{A}$ be the family of all multidigraphs such that all vertices are reachable from vertex 1 via a directed path, and let $A(w,z)$ be the corresponding bgf. There is a nice relation between $A(w,z)$ and the bgf $C(w,z)$ for connected undirected multigraphs, (2.10): If $A(w,z) = \Sigma_{n\geq 1}\, a_n(w) z^n/n!$, we have

$$\sum_{n\geq 1} a_n(w) e^{-n^2 w/2} \frac{z^n}{n!} = C(w,z)\,. \tag{29.11}$$

This can be proved by replacing z by $ze^{-w/2}$ and noting that $C(w, ze^{-w/2})$ is the

$d =$	0	1	2	3	4	5	6	7	8
$s_{1d} =$	$\frac{1}{2}$								
$s_{2d} =$	$\frac{17}{8}$	$\frac{13}{8}$	$\frac{1}{6}$						
$s_{3d} =$	$\frac{275}{12}$	$\frac{427}{12}$	$\frac{391}{24}$	$\frac{13}{6}$	$\frac{1}{24}$				
$s_{4d} =$	$\frac{26141}{64}$	$\frac{61231}{64}$	$\frac{51299}{64}$	$\frac{18473}{64}$	$\frac{6047}{144}$	$\frac{263}{144}$	$\frac{1}{120}$		
$s_{5d} =$	$\frac{1630711}{160}$	$\frac{1276481}{40}$	$\frac{3125933}{80}$	$\frac{2840093}{120}$	$\frac{3546283}{480}$	$\frac{6743}{6}$	$\frac{25307}{360}$	$\frac{43}{36}$	$\frac{1}{720}$

No reason why $S_r(z)$ should have the simple form (29.17) is apparent; this phenomenon cries out for explanation, if it is indeed true for all $r > 0$, and the explanation will probably lead to new theorems of interest. It can be shown that this conjecture is equivalent to the assertion that the sum of $(-1)^{\nu}\kappa/\nu!$, over all labelled, reduced, strongly connected multidigraphs of excess r, is zero; or in other words, if we choose a labelled, reduced, strongly connected multidigraph of excess r at random, with probabilities weighted in the natural way by the compensation factor κ, then the probability is $\frac{1}{2}$ that there will be an even number of vertices.

Is there a simple recurrence governing the leading coefficients $s_{10}, s_{20}, s_{30}, \ldots$, perhaps analogous to the relation we observed for ordinary connected components in (8.5)?