

How We Taught the Computer to Write in Persian

Behrooz Parhami

University of California, Santa Barbara

June 2024

چگونه فارسی نوشتن را به کامپیوتر آموختیم



بهر روز پرهامی

دانشگاه کالیفرنیا در سنن باربارا

June 2024

Writing the History of Computer-Generated Persian Script

Computers and Challenges of Writing in Persian:

Explorations at the Intersection of Culture and Technology

The Persian script has presented some difficulties, ever since printing presses were introduced in Iran in the 1600s. The appearance of typewriters more than a century ago created additional problems and the introduction of digital computers in the late 1960s added to the design challenges. These difficulties persisted, until high-resolution dot-matrix printers and display devices offered greater flexibility to font designers and the expansion of the computer market in the Middle East attracted investments on improving the Persian script for computers. Nevertheless, certain peculiarities of the Persian script have led to legibility and aesthetic quality issues that persist in many cases. In this paper, I enumerate some of the features of the Persian script that made it a poor match to implementation on modern technologies and review the challenges presented by, and some of the solutions proposed for, each new generation of printing and display devices. Before delving into the technical challenges and solutions, however, I discuss the sociocultural significance of the interplay between centuries-old Persian culture/language/script and modern technology, along with associated research problems. Interestingly, the same features that make legible and aesthetically pleasing Persian printing/displaying difficult also lead to challenges in automatic text recognition. I conclude with an overview of current state of the art and areas that still need further work.

Keywords -

Aesthetics
Arabic script
Computer display
Computer printing
Dot-matrix font
Font design
Keyboard legibility
Line printer
Moveable type
Persian script
Printing
Technology transfer
Typewriter

نگارش فارسی با کامپیوتر: مروری بر دشواری‌ها و راهکارها

بهروز پرهامی^۱

دانشکده مهندسی برق و کامپیوتر، دانشگاه کالیفرنیا در سنتا باربارا

این مقاله در پی تدوین تاریخچه‌ای کامل و مستند از فعالیت‌های گسترده و ابداعاتی است که نگارنده، همکاران دانشگاهی او، و متخصصان بخش‌های دولتی و خصوصی

صورت دیگری از این مقاله پیش از این در دو سخنرانی با عنوان "کامپیوتر و مشکلات نگارش فارسی" به زبان‌های فارسی در ۱۹ نوامبر و انگلیسی در ۲۰ نوامبر ۲۰۱۷ در سلسله سخنرانی‌های دوازده برنامه‌ی ایران‌شناسی دانشگاه یوسی‌ال‌ای عرضه شده است. از گرداننده سخنرانی‌ها، خانم دکتر نیره توحیدی، و سایر دست‌اندرکاران و حامیان آن برنامه صمیمانه سپاس‌گزارم. نقش نگارنده در پژوهش‌ها و راهکارهایی که در این مقاله به آنها اشاره می‌رود عمدتاً طبقه‌بندی مسائل، جمع‌بندی، و گزارش بوده است. مشکلات مورد بحث بزرگ‌تر از آن‌اند که یک نفر به تنهایی قادر به حلشان باشد. همکارانم در دانشگاه صنعتی شریف، آقایان دکتر فرهاد مودت و مهندس آرمن نهایتیان، در ابتدای این راه نقش مهمی ایفا کردند. همکاران دیگر دانشگاه‌های ایران و پژوهشگران و مهندسان سازمان‌های دولتی و بخش خصوصی و دانشجویان دوره کارشناسی ارشد دانشگاه صنعتی نیز در حل مسائل دست داشته‌اند. لازم به توضیح است که اشاره به مشکلات خط فارسی در این مقاله صورت‌های دیگر زبان فارسی، از جمله ذری، اردو، و پشتو را هم در بر می‌گیرد و مشکلات خط عربی هم عمدتاً شبیه فارسی است.

Behrooz Parhami, "Computers and Challenges of Writing in Persian," *Iran Namag*, Volume 4, Number 2 (Summer 2019), 3-32.

بهروز پرهامی <parhami@ece.ucsb.edu> (دانش‌آموخته دکتری علوم کامپیوتر، دانشگاه یوسی‌ال‌ای، ۱۹۷۳) استاد دانشگاه کالیفرنیا در سنتا باربارا و معاون پیشین کالج مهندسی آن است که در حال حاضر به تدریس و پژوهش در رشته معماری و سخت‌افزار کامپیوتر اشتغال دارد. او عضو برجسته انجمن بین‌المللی مهندسان برق و الکترونیک، انستیتوی مهندسی و تکنولوژی، و انجمن کامپیوتر بریتانیاست و جوایز متعددی دریافت کرده است. از ایشان ۶ کتاب و بیش از ۳۰۰ مقاله منتشر شده است. از ۱۹۷۴ تا ۱۹۸۶ در دانشگاه صنعتی شریف به مطالعه زبان و خط فارسی برای استفاده در کامپیوتر و مسائل انتقال تکنولوژی و استانداردسازی مشغول بوده و در تأسیس انجمن انفورماتیک ایران و نشریه آن انجمن با نام *گزارش کامپیوتر* مشارکت داشته است.



ایران نامگ، سال ۴، شماره ۲، تابستان ۲۰۱۹/۱۳۹۸

Publications on Persian Language and Computers

[041]

AUTOMATIC RECOGNITION OF PRINTED FARSI TEXTS

BEHROOZ FARHAMI and MAHMOOD TARAGHT
Tehran University of Technology

2. Farhani, Lower Bounds for Dot-Matrix Characters to Represent Farsi and Arabic Scripts

[045]

STANDARD FARSI INFORMATION INTERCHANGE CODE AND KEYBOARD LAYOUT: A UNIFIED PROPOSAL

BEHROOZ FARHAMI

Department of Mathematics & Computer Science, Sharif University of Technology, PB 246, Tehran, Iran

Page 1 of 8

[111]

On Lower Bounds for the Dimensions of Dot-Matrix Characters to Represent Farsi and Arabic Scripts

Behrooz Farhami
Department of Electrical and Computer Engineering
University of California
Santa Barbara, CA 93106-9560, USA

Abstract: Lower bounds for the dimensions of dot-matrix characters to represent Farsi and Arabic scripts are derived. It is shown that for a given number of dots, there is a minimum number of dots that must be used to represent a character.

Copyright © 1983, IEEE. All rights reserved.

LANGUAGE-DEPENDENT CONSIDERATIONS FOR COMPUTER APPLICATIONS IN FARSI AND ARABIC SPEAKING COUNTRIES

B. Farhami

Department of Mathematics and Computer Science,
Sharif University of Technology, Tehran, Iran

[023]

ABSTRACT. A prerequisite to economic viability and social acceptance of computer-based systems, particularly in developing countries, is their ability to accept and present information in the natural language of the user community. This paper identifies problems relating to the use of Farsi and Arabic languages in computer-based systems and outlines known approaches to their solution. The problems discussed fall into four main categories: (1) Information representation and standardized information interchange codes; (2) Input or information through devices such as keyboards and optical readers; (3) Processing functions such as sorting, text formatting and data compression; and (4) Output of information on hard-copy devices and various types of displays.

Keywords: Character recognition; Computer applications; Computer input and output; Computer peripheral equipment; Data processing; Display systems; Name factors; Printers; Standards; Text editing.

INTRODUCTION

Computers are information processing machines and much of the information we deal with in our everyday lives is generated, maintained, and used in a natural language. Furthermore, for computers to be useful, they must interact with human beings and the most convenient way of doing this, at least as far as ordinary users are concerned, is by utilizing a natural language as the medium of communication. It follows that the ability of computers to deal with information presented in a natural language is essential for their successful utilization in most environments.

Even though language considerations are not critical to activities such as programming, there are clearly instances where information storage (e.g., in the case of names and addresses), manipulation (e.g., for processing), and presentation (e.g., for display) of information can be most effectively handled in the users' native language. It is, therefore, quite natural that for a number of years, we in Iran have been concerned with language-related problems

considerations only to Arabic and Urdu as well.

REQUIREMENTS OF INFORMATION

The recording of information on computer storage media and its translation from one place to another (as they circulate on the same media) requires some form of digital encoding. The advantages of using standard codes for these purposes are well known and have resulted in the adoption of national and international standards for various natural languages. National standards of this type must be designed according to specific guidelines set forth by the International Standards Organization (ISO, 1973), if they are to be mutually compatible.

Following relative activity in various application areas, the need for a standard representation of Farsi symbols was felt in Iran (Dezouzi and Farhani, 1977). In February of 1979, the Iranian Farsi and Subject Organization invited a number of computer and language experts to form the "Study Committee

[033]

ON THE USE OF FARSI AND ARABIC LANGUAGES IN COMPUTER-BASED INFORMATION SYSTEMS

BEHROOZ FARHAMI

Arya-Wehr University of Technology
Tehran, Iran

[032]

SPECIALLY WEIGHTED DOT-MATRIX FARSI AND ARABIC NUMERALS

BEHROOZ FARHAMI

Arya-Wehr University of Technology
Tehran, Iran

Abstract: Representation of digits by specially weighted numerals to a set of standard information processing systems, with the clear objective of being able to handle them in a uniform manner, is discussed. A general framework for the design of specially weighted numerals is established and several examples are given to illustrate the concept. It is shown that Farsi and Arabic numerals can be represented by specially weighted numerals and that such numerals are suitable for automatic processing.

Keywords: Arabic numerals; character set design; Farsi numerals; numeric output; optical readers; specially weighted numerals; text editing; text processing; text representation; text storage; text transmission.

[024]

INFORMATICS IN IRAN: PROBLEMS AND PROSPECTS

BEHROOZ FARHAMI and MAHMOOD TARAGHT
Tehran University of Technology
Tehran, Iran

Abstract: This is a preliminary report on the state of informatics in Iran. It distinguishes the case of Iran from those of many other developing countries in its dedication to modernization and use of computers with very little budgetary restriction. The Government's commitment to modernization and its clearly stated policy of computerization have created many unique problems in Iran which intensify the difficulties faced by other developing countries. In this report, the major problem areas are enumerated and analyzed, and a few measures for coping with them are proposed.

1. INTRODUCTION

Some musical instruments such as pianos and accordions do not produce the undesirable harmonics and, therefore, they can be pleasing to the ear even when played by beginners. While certain others, notably string instruments, do produce these harmonics and can have pleasing results even with semi-skilled players. The ability of beginners to produce pleasing effects has also been the case with more complex and more recently with computers.

Computers are easily played instruments because of their tremendous power, speed, and accuracy. They can produce interesting output well justified, nicely formatted, carefully analyzed and sorted. While in many parts of the world, highly trained ears and eyes have made it difficult for non-expert musicians and movie makers to impress the public, most recipients of computer services are ignorant of the state of computer art, and pleasing them is still not so difficult. This fact is thoroughly exploited by non-skilled computer players around the world, particularly in developing countries where the users of computer services have minimal knowledge of computing. Influenced by

[111]

INFORMATION PROCESSING II, B. GILCHRIST, EDITOR
© IFFP, NORTH-HOLLAND PUBLISHING COMPANY (1977)

COMPUTERS AND THE FARSI LANGUAGE—A SURVEY OF PROBLEM AREAS

BEHROOZ FARHAMI and FARHAD MAVADDAT
Arya-Wehr University of Technology
Tehran, Iran

As users of computer systems advanced by the west, we face various problems brought about by the fact that the Farsi language and the field of computing. It is felt that a satisfactory solution of these problems is essential to the development of our country in this century. We review these problems by dividing them into five categories: namely, education, cost, staff, hardware, and programming. For each category, we present an introduction to the problem in its particular area and discuss past approaches and promising techniques.

1. INTRODUCTION

The application of computers to Iran started in 1962 with the installation of an IBM 1620. Since then, the number of computers has increased roughly by a factor of ten every five years. The technology of computers in Iran brought about many problems. The most important being a lack of knowledge of the use of computers, a lack of staff, and a lack of hardware. These problems are discussed in this paper. The first two problems are shared by other developing countries, the third one is more specific to Iran.

The purpose of this paper is to survey problem areas in the field of computing which are related to the Farsi language. Some though computers and software manufacturers have attempted to alleviate some of these problems, we do not consider the proposed solutions satisfactory. This is due to the fact that the solutions in such development efforts and the associated cost are not acceptable. In our opinion, a brief introduction to the Farsi language is necessary. Farsi is a member of the Indo-European family and has varying numbers and helps to adjust to the same word through the same time. In our opinion, the Farsi language is not so different from other languages as it is often claimed. In all Farsi computer systems, the use of the Farsi language is not so different from other languages as it is often claimed. In all Farsi computer systems, the use of the Farsi language is not so different from other languages as it is often claimed. In all Farsi computer systems, the use of the Farsi language is not so different from other languages as it is often claimed.

TABLE I
Correspondence with representation of
Alphabetic Symbols

Alphabetic Symbol	Representation
ا	0
ب	1
پ	2
ت	3
ث	4
ج	5
چ	6
ح	7
خ	8
د	9
ذ	10
ر	11
ز	12
س	13
ش	14
ص	15
ض	16
ط	17
ظ	18
ع	19
غ	20
ف	21
ق	22
ک	23
گ	24
ن	25
ی	26
و	27
ه	28
و	29
ی	30
و	31
ی	32
و	33
ی	34
و	35
ی	36
و	37
ی	38
و	39
ی	40
و	41
ی	42
و	43
ی	44
و	45
ی	46
و	47
ی	48
و	49
ی	50
و	51
ی	52
و	53
ی	54
و	55
ی	56
و	57
ی	58
و	59
ی	60
و	61
ی	62
و	63
ی	64
و	65
ی	66
و	67
ی	68
و	69
ی	70
و	71
ی	72
و	73
ی	74
و	75
ی	76
و	77
ی	78
و	79
ی	80
و	81
ی	82
و	83
ی	84
و	85
ی	86
و	87
ی	88
و	89
ی	90
و	91
ی	92
و	93
ی	94
و	95
ی	96
و	97
ی	98
و	99
ی	100

2. COMPUTER SCIENCE EDUCATION

Computer education in Iran, even at the elementary level, is considered a luxury. The number of computers and the staff available for teaching is very limited.

On-Line Access to My Pertinent Publications

Three-digit numbers in square brackets refer to entries in B. Parhami's Publications list:

<http://www.ece.ucsb.edu/~parhami/publications.htm>

[111] B. Parhami, "On Lower Bounds for the Dimensions of Dot-Matrix Characters to Represent Farsi and Arabic Scripts," *Proc. 1st Annual CSI Computer Conf.*, Tehran, Iran, December 1995, pp. 125-130.

[045] B. Parhami, "Standard Farsi Information Interchange Code and Keyboard Layout: A Unified Proposal," *J. Institution of Electrical and Telecommunications Engineers*, Vol. 30, pp. 179-183, 1984.

[041] B. Parhami and M. Taraghi, "Automatic Recognition of Printed Farsi Texts," *Pattern Recognition*, Vol. 14, Nos. 1-6, pp. 395-403, 1981.

[038] B. Parhami, "Language-Dependent Considerations for Computer Applications in Farsi and Arabic Speaking Countries," *System Approach for Development (Proc. IFAC Conf.)*, 1981, pp. 507-513.

[037] B. Parhami and M. Taraghi, "Automatic Recognition of Printed Farsi Texts" (summary), *Proc. Conf. Pattern Recognition*, Oxford, England, January 1980.

[033] B. Parhami, "On the Use of Farsi and Arabic Languages in Computer-Based Information Systems," *Proc. Symp. Linguistic Implications of Computer-Based Information Systems*, November 1978.

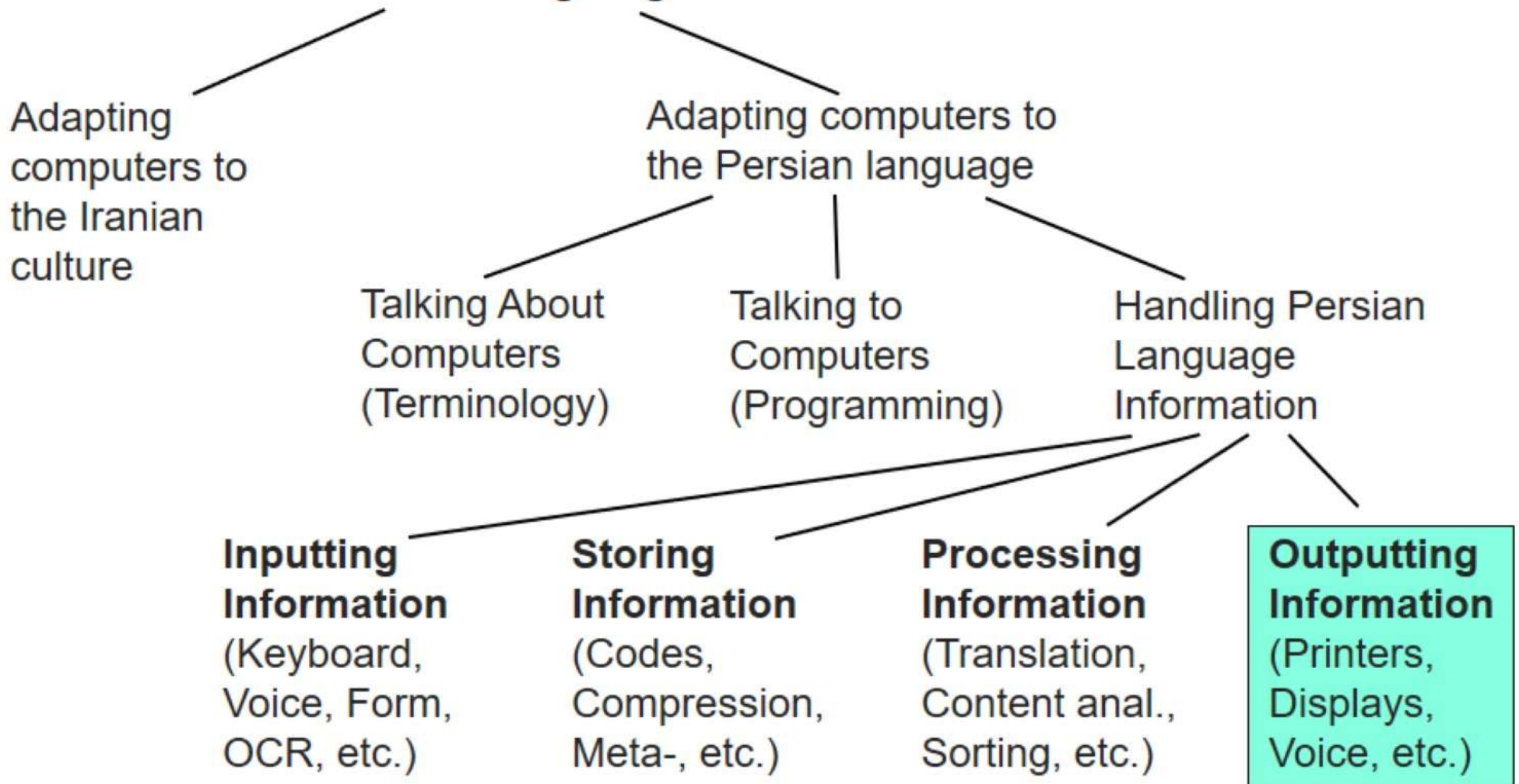
[032] B. Parhami, "Optically Weighted Dot-Matrix Farsi and Arabic Numerals," *Information Technology 78 (Proc. 3rd Jerusalem Conf.)*, August 1978, pp. 207-210.

[023] B. Parhami and F. Mavaddat, "Computers and the Farsi Language: A Survey of Problem Areas," *Information Processing 77 (Proc. IFIP World Congress)*, 1977, pp. 673-676.

[. . .] Parhami, B., "Impact of Farsi Language on Computing in Iran," *Mideast Computer*, Vol. 1, pp. 6-7, September 1978. [Not part of my academic publications; link to be supplied here]

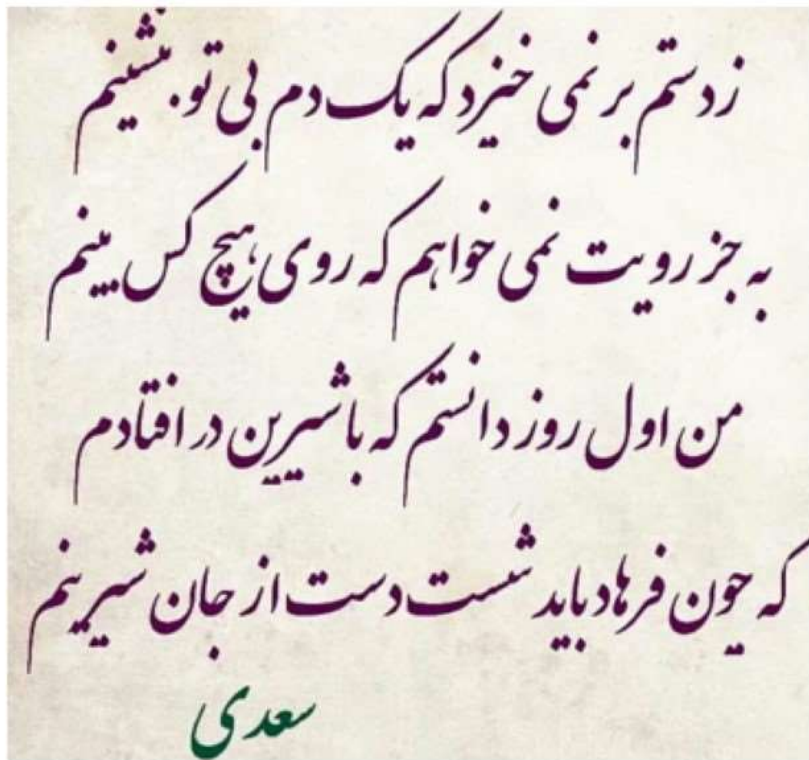
Computers and the Persian Language: Technology Transplantation (not Transfer)

Culture and Language Issues



Let's Start at the Very Beginning: A Very Good Place to Start

Nastaliq script was used
by scribes and
calligraphic artists



Text by
unknown
scribe



Art by
Farokh
Mahjoubi

Rules for Beautiful Writing: Nasta'liq and Naskh Scripts

مرتبہ قواعد و قوانین خط نستعلیق (طیلسن)

تجربہ درمیانِ شمش و طمع و ف
ق ک گ ل م ن و ہ لاء مہم ہے ہی سے

تعمیر الہیہ و تزیین ہر حرف و کلمہ - الامام ابو جعفر محمد بن یوسف الخزاز نقی

فانفت فح قدر فرس شس فص فط فع
فق قک قل قم فن فو فہ فہ فلا فس فی

ا ب د ر نین
ص ط ع ک
ل م ن ی و

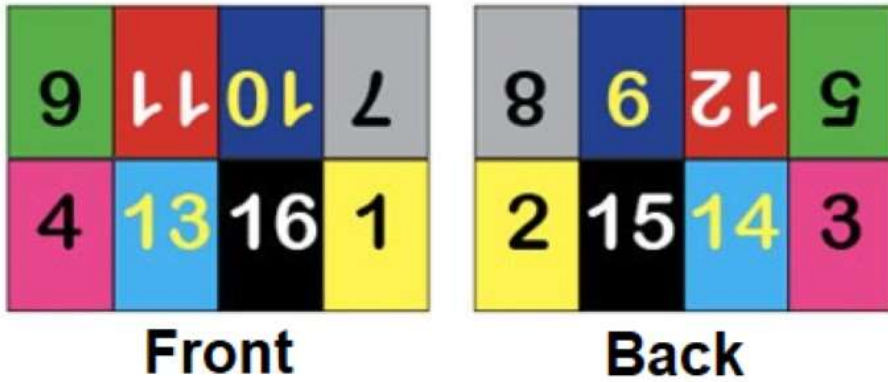
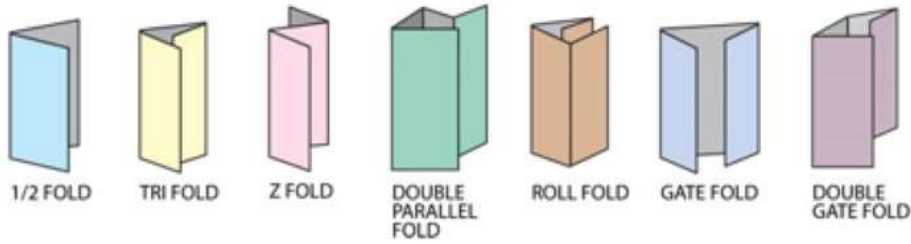
fi ffi لا ← ا + ل U+0627 U+0644

Final shape Initial shape ل ← ا + ل ← ا + ل U+0627 U+0644

The Printing Press and Movable Font



Image Transfer to Paper and Binding



Publishing in Persian: A Global View



Publishing in Persian language

in Iran, Afghanistan, Tajikistan, Uzbekistan, Europe
and United States

Farid Moradi, Laetitia Nanquette,
Masoud Hosseinipour, Ali Amiri, Dilshad
Rakhimov, B. Biniaz



http://www.alliance-editeurs.org/IMG/pdf/publishing_in_persian_language-2.pdf

Movable Persian Fonts Developed



Nov. 2017



Fifty Years of Poor Penmanship



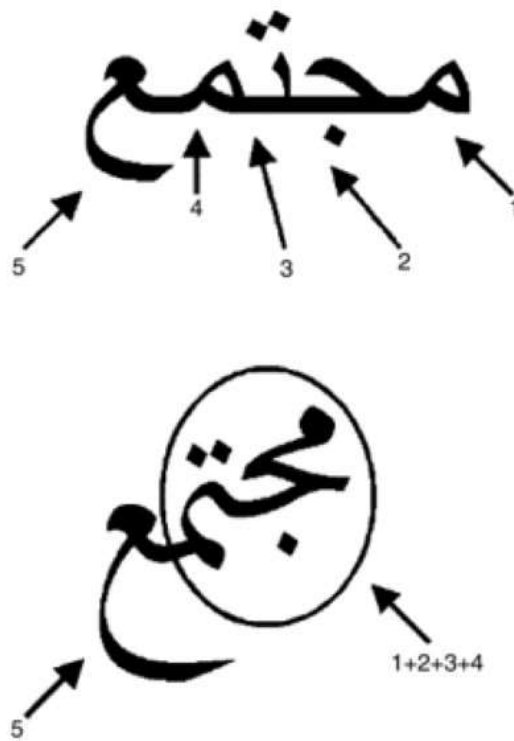
Slide 15

Early Challenges with Print Typefaces: Decomposition and Added Uniformity

Naskh script was more suitable for adaptation to printing press, which came to Iran in the 1600s

لحظه بخط

قاشق
و
چنگال



تخت جمشید

تخت جمشید یا پارسه (یا پرسپولیس، پرسه پلیس، هزارستون، صدستون و یا چهل منار) نام یکی از شهرهای باستانی ایران است که طی سالیان پیوسته، پایتخت باشکوه و تشریفاتی پادشاهی ایران در زمان امپراتوری هخامنشیان بوده است. در این شهر باستانی کاخی به نام تخت جمشید وجود دارد که در دوران زمامداری داریوش

چو شهر عشق من شهری ندیدم
ز نادانی بسی غربت کشیدم
هر آوازی که در عالم شنیدم
چو ماری بی پرو یا میخزیدم
چو گل بی حلق و بی لب می چشیدم
که من محنت سرائی آفریدم
از آنجا آمدن هم می رمیدم
که من نزدیک، چون جبل الوری دم
فسون و عشوه او را خریدم

سفر کردم، بهر شهری دویدم
ندانستم ز اول قدر آن شهر
بغیر از عشق آواز دهل بود
در آن عالم که جان بودم به تنها
از آن باده که لطف و خنده بخشد
ندا آمد ز عشق: ای جان سفر کن
چنان کاکنون ز رفتن میگریزم
بگفت: ای جان برو هر جا که خواهی
فسون کرد و مراس عشوه ها داد

ز راهم برد و آنجا هم بره کرد
کز آن ره من نرفتم من رهیدم

مولوی (مولانا جلال الدین محمد بلخی رومی)

Persian vs. Arabic Printing: Problems are Quite Similar

العربية
فارسی

Arial Bold:

ليبيا: الامم المتحدة تدعو لوقف
الصراع تخفيفا لمعاناة الناس

Arabic Transparent Bold:

ليبيا: الامم المتحدة تدعو لوقف
الصراع تخفيفا لمعاناة الناس

Times New Roman Bold:

ليبيا: الامم المتحدة تدعو لوقف
الصراع تخفيفا لمعاناة الناس

Simplified Arabic Bold:

ليبيا: الامم المتحدة تدعو لوقف
الصراع تخفيفا لمعاناة الناس

Tahoma Bold:

ليبيا: الامم المتحدة تدعو لوقف
الصراع تخفيفا لمعاناة الناس

بلادين balā-dīdan [ع. - ف.]
(مصل. ١) - رنج دیدن. ٢ - بمصیبت
دچار شدن.

بلاديه b.-dīda(-e) (ص.مف.) ١ -
رنجیده. ٢ - دچار مصیبت شده.

بلاذر balāzor [= بلادر = بلادور]
(١.) (گیا.) ١ - بلادر (م.ه.)

بلاز balār [= بلال] (١.) (گیا.)
آذر بویه (م.ه.)، اشنان (م.ه.)

بلار ballār [= بلور] (١.) ← بلور.

بلارج balāraġ (١.) (جاز.) لكلك
← ٢ - لكلك

الله

ل + ا = لا
نرنا نرنا نرنا

Some Properties of the Persian Script: Greater Variety in Letter Shapes and Sizes

The Latin alphabet

a b c d e f g h i j k l m n o p q r s t u v w x y z + / # % \$ & (
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

The quick brown fox jumped over the fence. Arial 24

The quick brown fox jumped over the fence. Times N R 24

The quick brown fox jumped over the fence. Courier 20

The Persian alphabet



Fixed-width characters

کار پاکان را قیاس از خود مگیر	گر چه باشد در نوشتن شیر، شیر
آن یکی شیر است اندر بادیه	آن دگر شیر است اندر بادیه
آن یکی شیر است کادم می خورد	وآن دگر شیر است کادم می خورد

Variable-width characters

کار پاکان را قیاس از خود مگیر	گر چه باشد در نوشتن شیر، شیر
آن یکی شیر است اندر بادیه	آن دگر شیر است اندر بادیه
آن یکی شیر است کادم می خورد	وآن دگر شیر است کادم می خورد

Persian Typefaces Developed into Modern Versions Now in Use



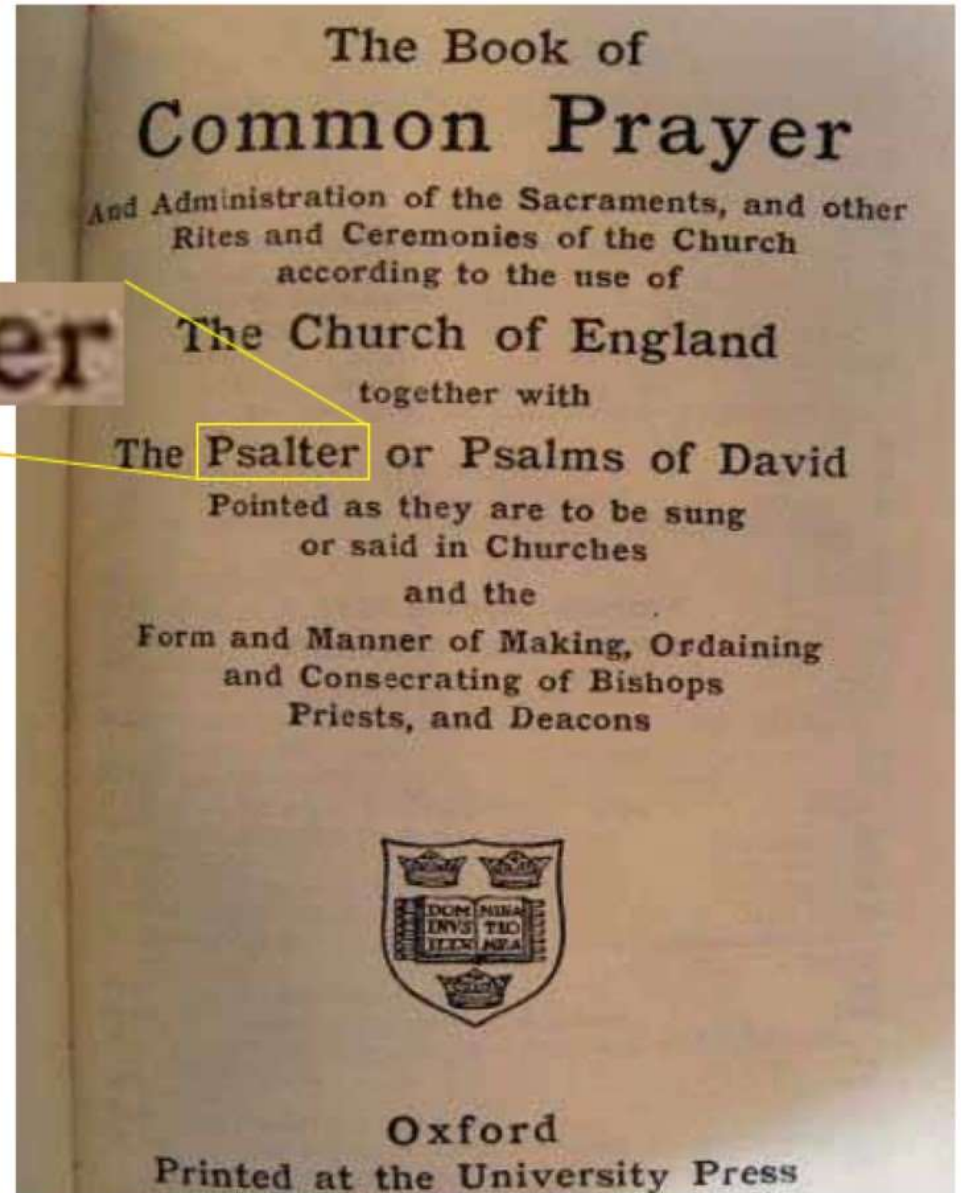
Early Computer Typesetting Challenges

Manually typeset text had:

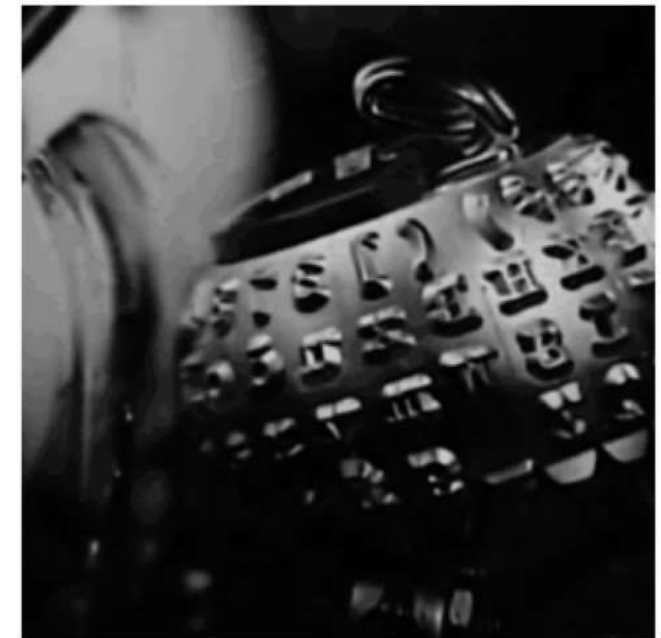
- Alignment problems
- Spacing irregularities
- Broken fonts
- Ink imperfections
- Nonuniform shapes

Early computer-typeset
Texts were disliked because
they were just too perfect

So, a little randomness was
introduced in font shapes
and alignment



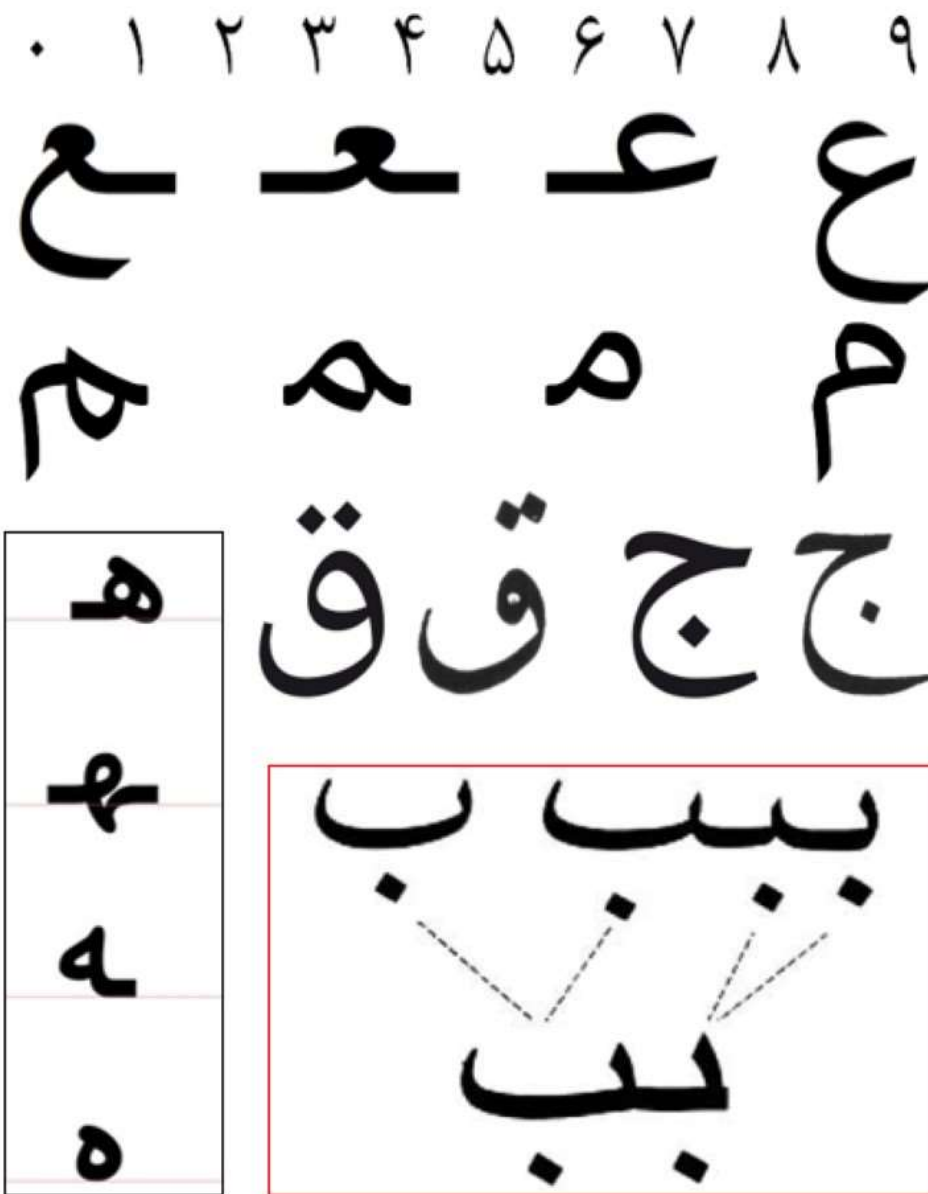
Persian Typewriters: Mechanical/Electric



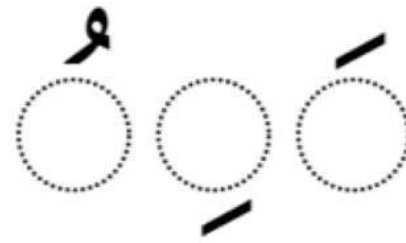
From Print to Typewriter Typefaces

Persian Alphabet / الفبای فارسی

Detached	Initial	Medial	Final	Roman	Name	Detached	Initial	Medial	Final	Roman	Name
ا	ا	ا	ا	á	alef	ص	ص	ص	ص	ş	sád
ب	ب	ب	ب	b	be	ض	ض	ض	ض	đ	zád
پ	پ	پ	پ	p	pe	ط	ط	ط	ط	t	tá
ت	ت	ت	ت	t	te	ظ	ظ	ظ	ظ	z	zá
ث	ث	ث	ث	th	se	ع	ع	ع	ع	'	ayn
ج	ج	ج	ج	j	jim	غ	غ	غ	غ	gh	ghayn
چ	چ	چ	چ	ch	che	ف	ف	ف	ف	f	fe
ح	ح	ح	ح	h	he	ق	ق	ق	ق	q	qáf
خ	خ	خ	خ	kh	khe	ك	ك	ك	ك	k	káf
د	د	د	د	d	dál	گ	گ	گ	گ	g	gáf
ذ	ذ	ذ	ذ	dh	zál	ل	ل	ل	ل	l	lám
ر	ر	ر	ر	r	re	م	م	م	م	m	mím
ز	ز	ز	ز	z	ze	ن	ن	ن	ن	n	nún
ژ	ژ	ژ	ژ	zh	zhe	و	و	و	و	v/ú	váv
س	س	س	س	s	sin	ه	ه	ه	ه	h	he
ش	ش	ش	ش	sh	shin	ی	ی	ی	ی	y/í	ye



Persian Typewriters: Memory and Intelligence



ک
کم
کمت
کمتر

حرف "ع" در چاپ فارسی چهار گونه دارد:
گونه ی اول، یا "ع"
گونه ی وسط، یا "ع"
گونه ی اخر، یا "ع"
گونه ی تنها، یا "ع"

حرف "ع" در چاپ فارسی چهار گونه دارد:
گونه ی اول، یا "ع"
گونه ی وسط، یا "ع"
گونه ی اخر، یا "ع"
گونه ی تنها، یا "ع"

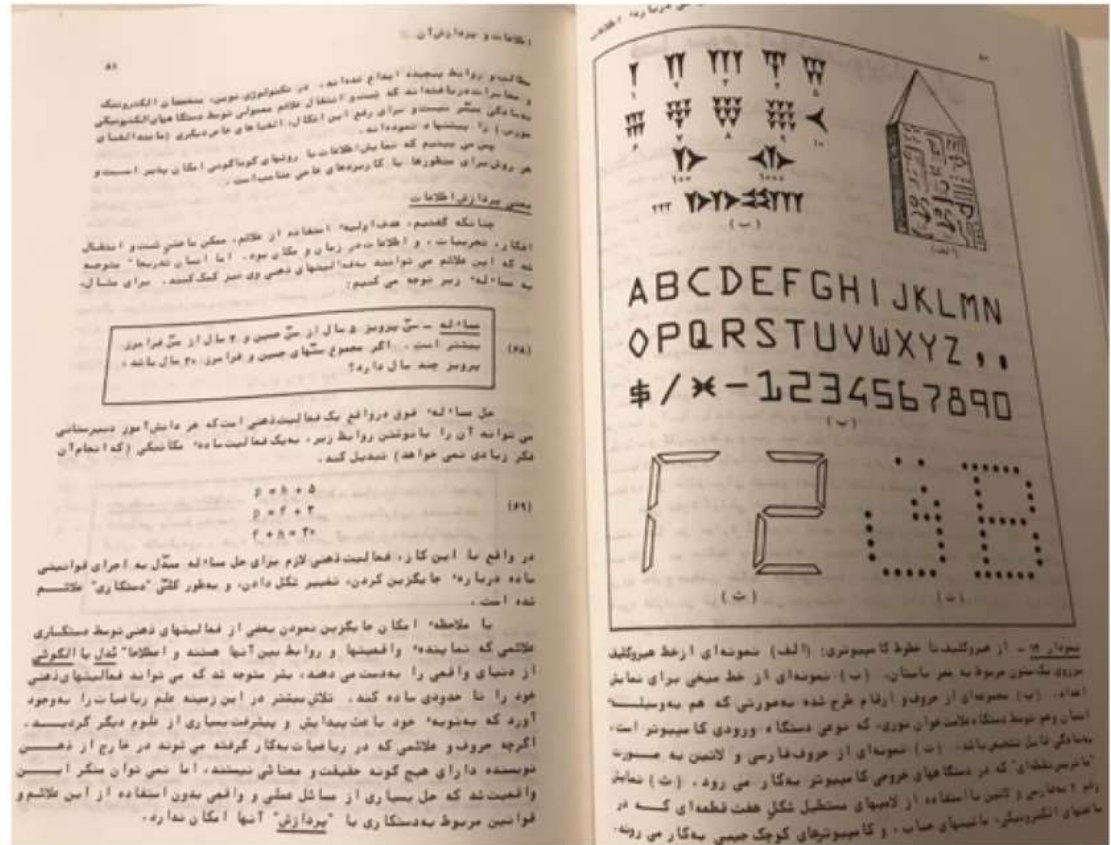
My 1984 Book Typed on IBM Selectric



آشنائی با کامپیوتر

تألیف دکتر بهروز پرهامی

چهارم و پنجم ۱۳۵۷ هجری قمری - انتشارات دانشگاه تهران



کامپیوتر در ایران

کاربرد کامپیوتر در ایران از سال ۱۳۴۱ خورشیدی، کوچک "آی بی ام ۱۶۲۰" آغاز گردید. برطبق آمار موجود ایران در سال ۱۳۵۰ به حدود ۱۰۰ دستگاه رسید، در سال ۵۶ به روایتی بالغ بر ۷۰۰ دستگاه گردید، و از آن پس تقریباً

Keyboard Layout and Exchange Code

!	"	#	٢	٪		١	()	=	÷	x	?	/
١	٢	٣	٤	٥	٦	٧	٨	٩	٠	-	+	—	ء
ض	ص	ث	ق	ف	غ	ع	ه	ح	خ	ج	ب	ا	DEL
ش	س	ي	ب	ل	ت	ن	م	ك	پ	د	ذ	ر	ء
ظ	ط	ز	ر	ذ	س	د	و	ز	/	*	١	٢	٣

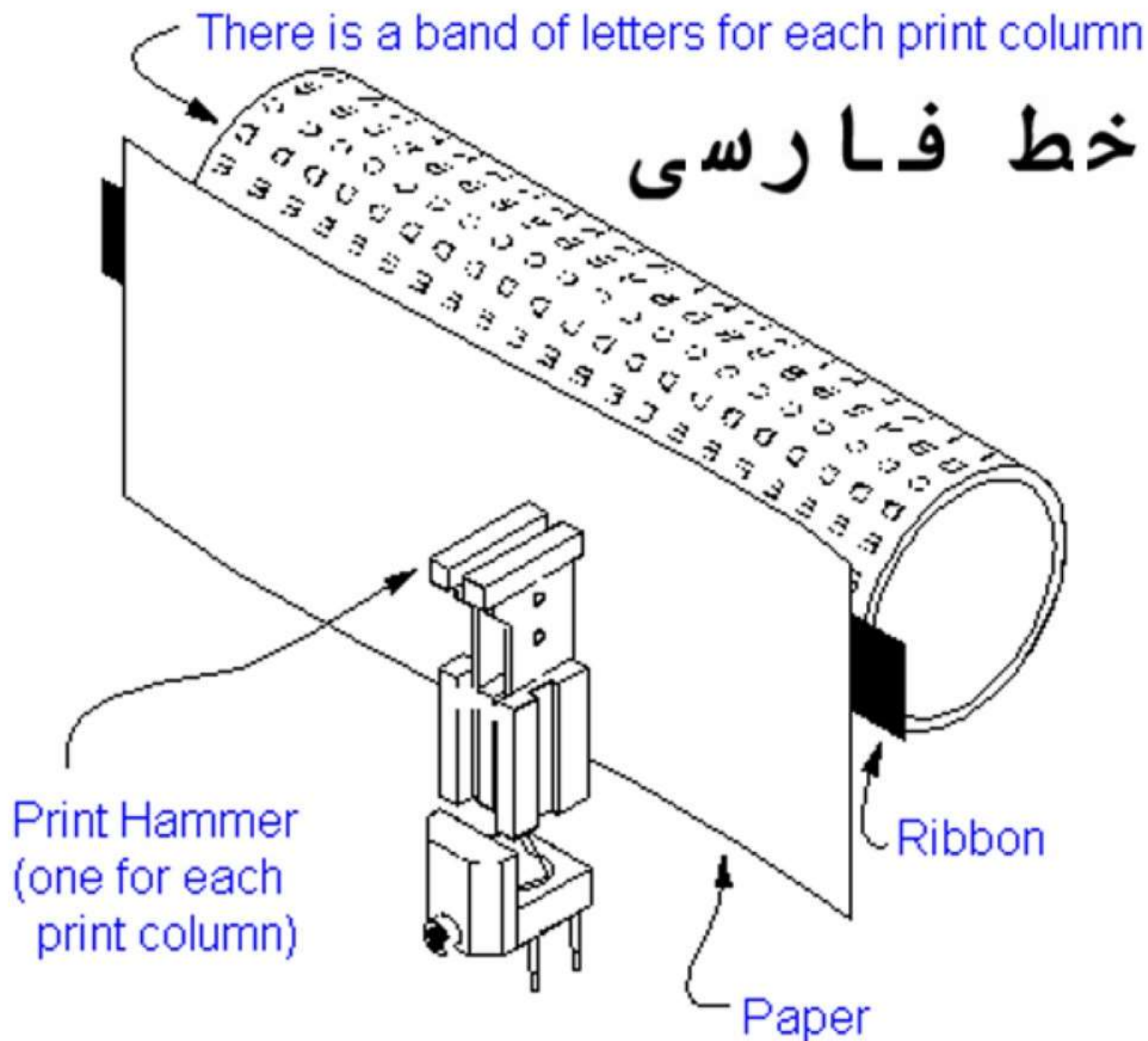
!	"	#	\$	%		/	()	=	-	\	?	
1	2	3	4	5	6	7	8	9	0	-	+	—	
Q	W	E	R	T	Y	U	I	O	P	<	>	DEL	
										&	@	—	
A	S	D	F	G	H	J	K	L	{	}	^		
										[]	^	
Z	X	C	V	B	N	M	,	:	;	:	*		

		G1						
		Column No.						
		2	3	4	5	6	7	
Row No.	0	SP	٠	x	٧	—	ل	
	1	!	١	٢	خ	ض	ز	
	2	"	٢	٣	خ	ط	م	
	3	#	٣	٤	د	ط	م	
	4	٢	٤	١	ذ	ظ	ن	
	5	٪	٥	٤	ر	ظ	ذ	
	6	÷	٦	ذ	ز	ع	و	
	7	'	٧	:	ز	ء	ه	
	8	(٨	:	ء	غ	ه	
	9)	٩	ز	ء	غ	ي	
	10	*	:	ز	ص	ف	ي	
	11	+	:	ج	'	ق	ء	
	12	,	«	ج	'	ق		
	13	-	=	ج	'	ك	ء	
	14	.	»	ج	'	ك	—	
	15	/	?	ح	—	لا	DEL	

		G0							
		Column No.							
		0	1	2	3	4	5	6	7
CONTROL CHARACTERS	0	@	P	`	p				
	1	A	Q	a	q				
	2	B	R	b	r				
	3	C	S	c	s				
	4	D	T	d	t				
	5	E	U	e	u				
	6	F	V	f	v				
	7	G	W	g	w				
	8	H	X	h	x				
	9	I	Y	i	y				
	10	J	Z	j	z				
	11	K	[k	{				
	12	L	\	l					
	13	M]	m	}				
	14	N	^	n	~				
	15	O	_	o					

ISO Code
(International Reference Version, ISO-IRV)

Early Persian Drum Printers



Drum line-printers:
Could print an entire
line at once (during
one drum revolution)

Gap between
hammers and letters
on the drum did not
allow good
connectivity of letters

One solution:
Print vertically

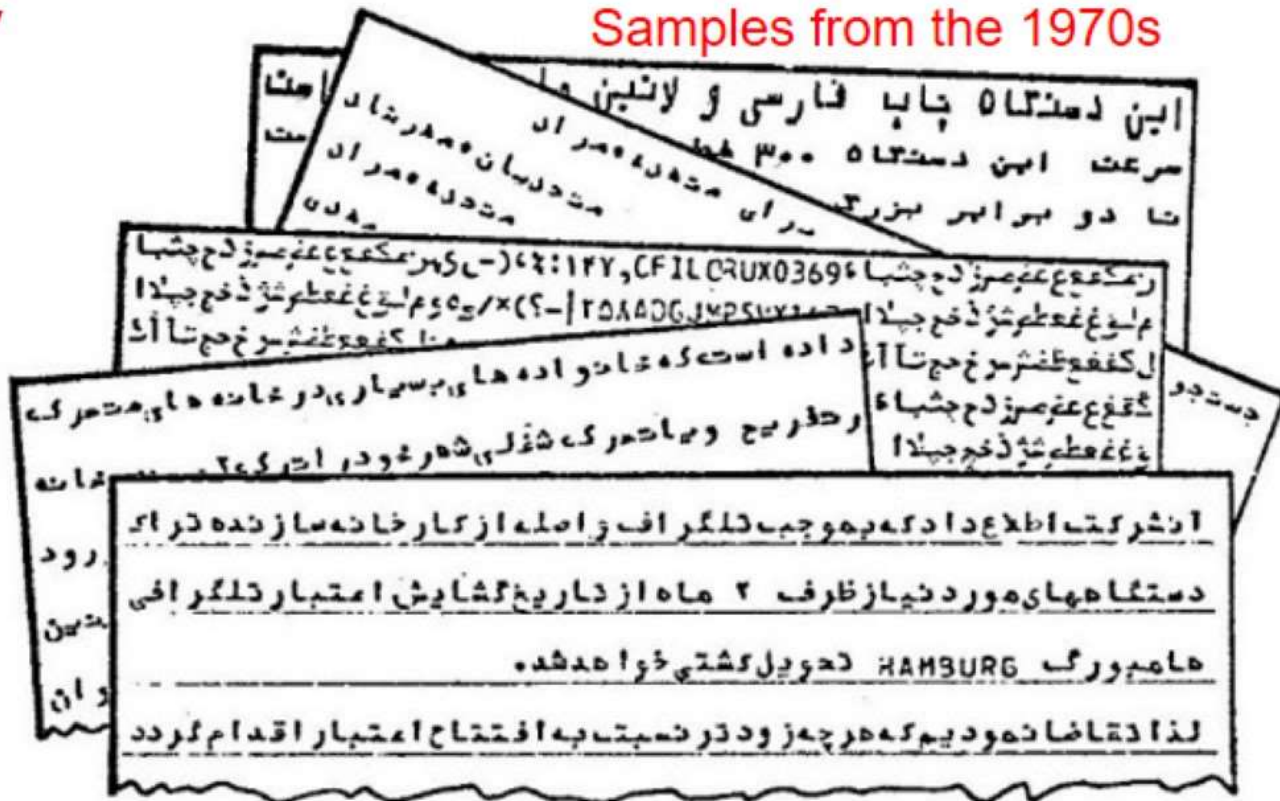
Samples from Early Persian Printers

Smudge problem of hammer-type printers was actually a plus, as it created the illusion of connectivity

However, it wasn't so good in other respects, because small features such as the loop in letters such as **قاف** and **میم** tended to disappear



Samples from the 1970s

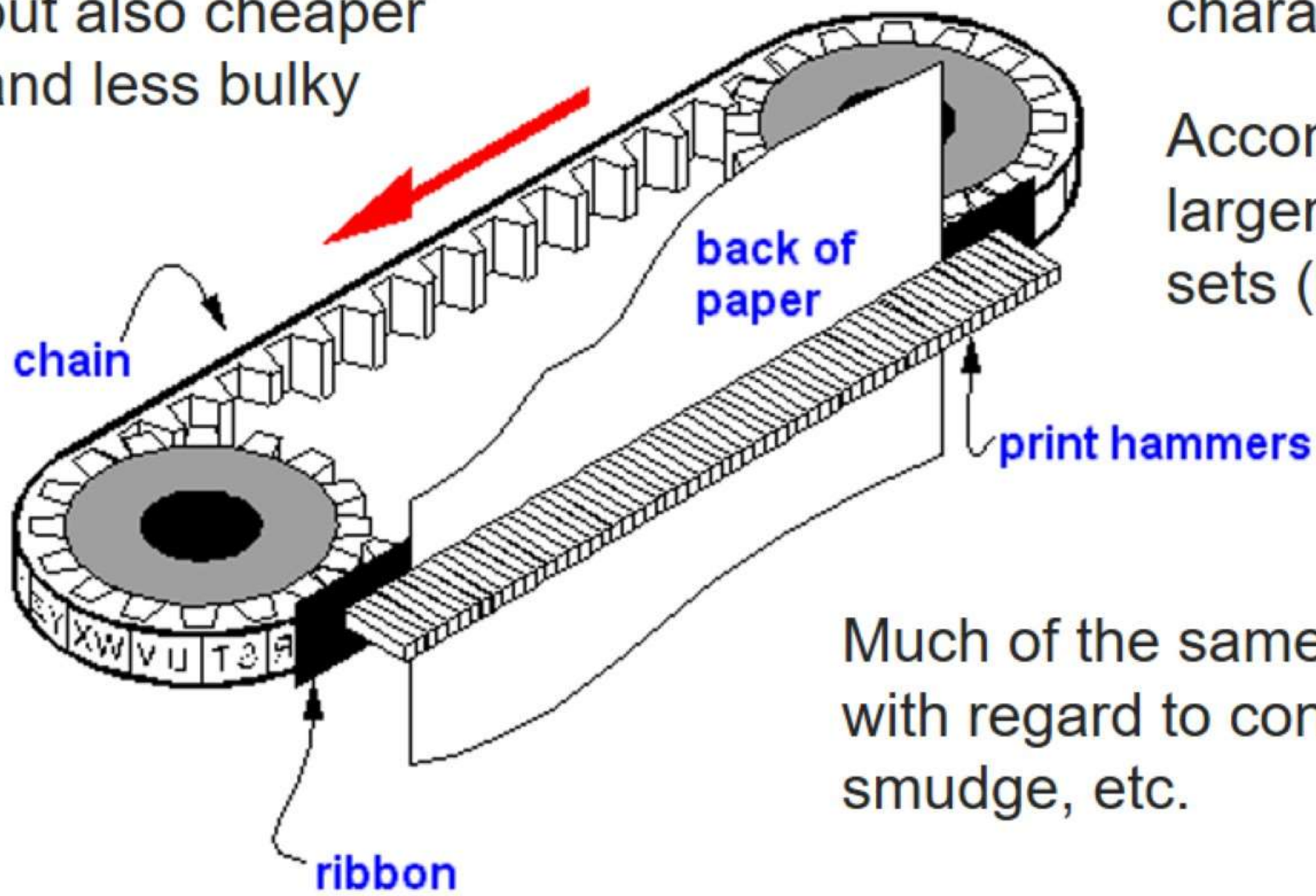


Early Persian Chain Printers

Slower than drum printers,
but also cheaper
and less bulky

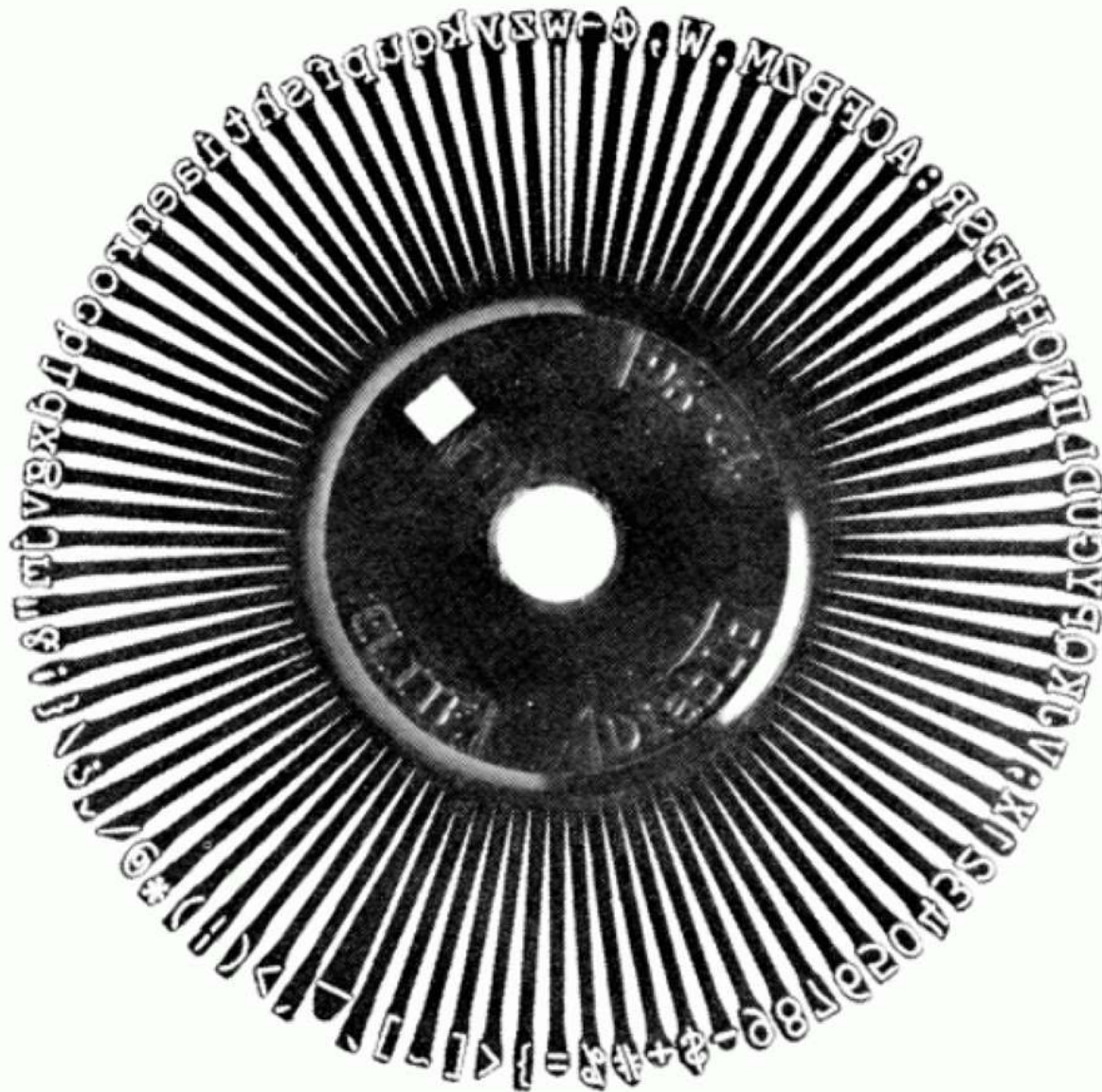
Fixed-width
characters

Accommodates
larger character
sets (bilingual?)



Much of the same problems
with regard to connectivity,
smudge, etc.

Early Persian Daisy-Chain Printers

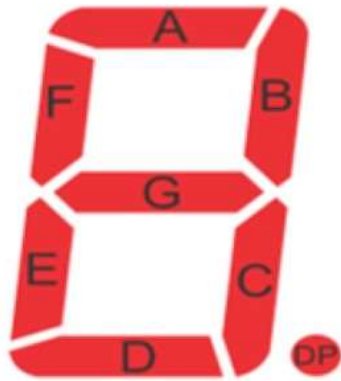


Even slower, but also cheaper and more compact

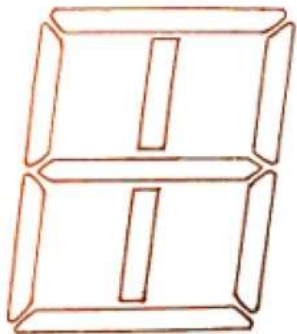
Fixed-width characters

Does not support multilingual printing or multiple font sizes

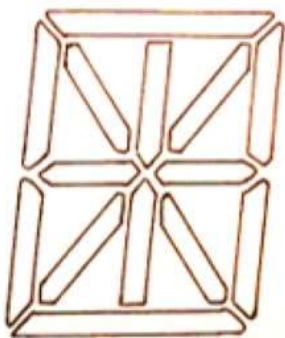
Lin-Segment Displays: Latin Script



7-segment display
for calculators
and other
numerical apps



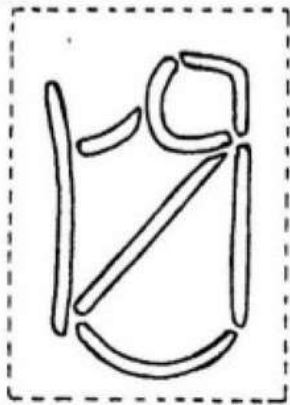
Two kinds of
line-segment
alphanumeric
displays



Okay for
uppercase
letters, but
problematic
for all else



Lin-Segment Displays: Persian Script



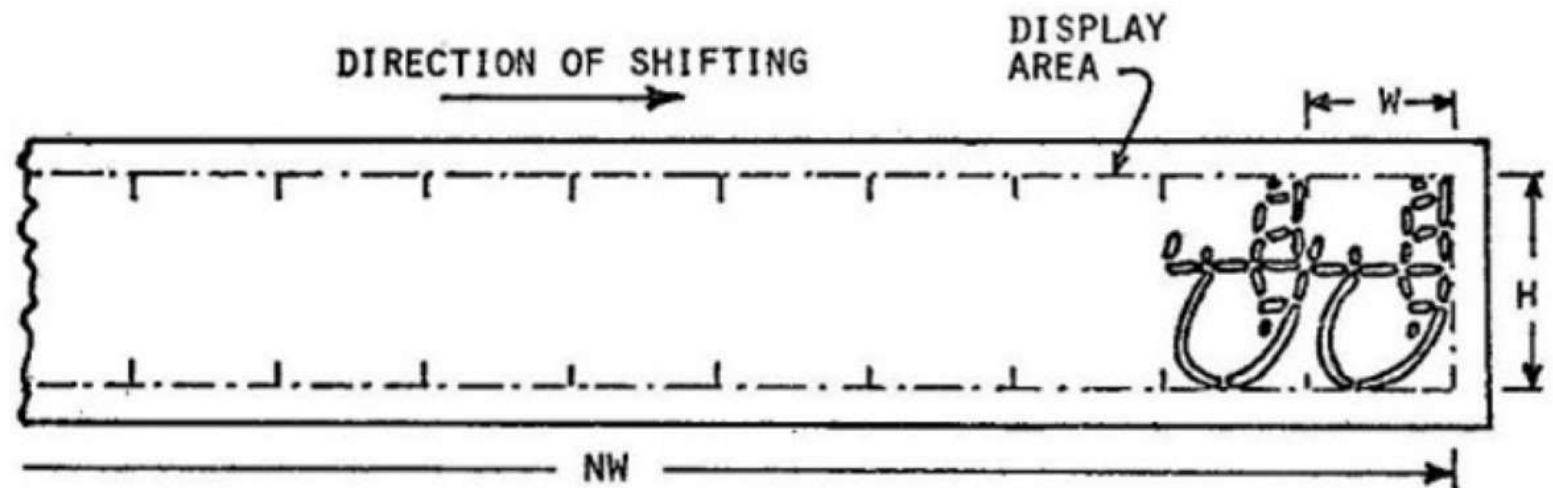
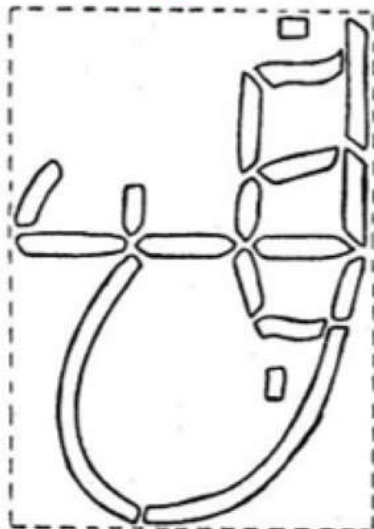
7 segments



Seven-segment display for Persian digits feasible, but ...

**At least 18 segments are needed for displaying letters
Even then, alphanumeric symbols cannot be combined
Special characters also present problems**

18 segments



Dot-Matrix Printers to the Rescue



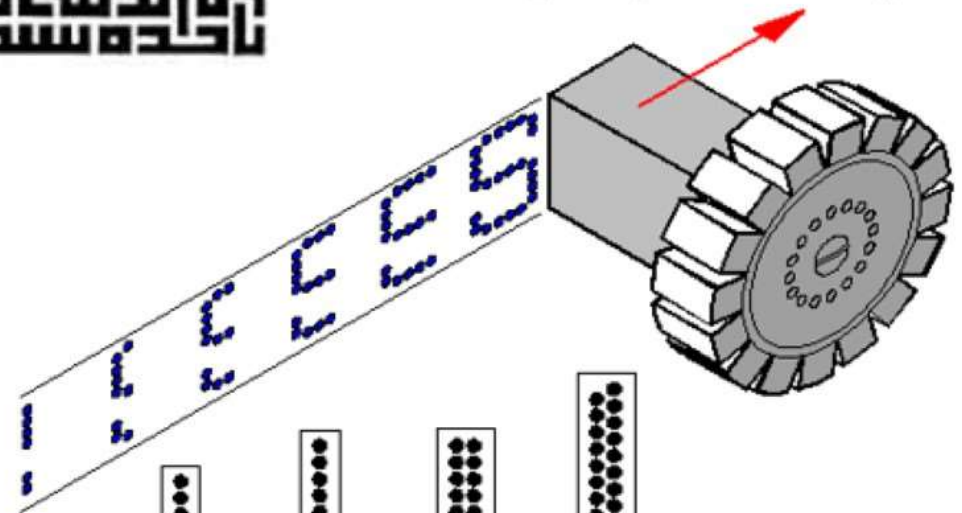
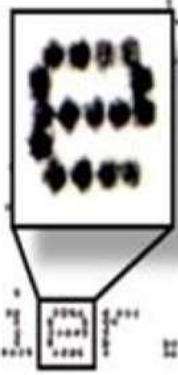
Kufi Script



PINEL FONT

A B C D E F G H I
 J K L M N O P Q R
 S T U V W X Y Z
 1 2 3 4 5 6 7 8 9 0
 ? ! @ # \$ % & * + = - _ #
 ^ / \ | . : ; ' () \$

ystem where a
 ld allow us t
 mercial supplier.



7 Pin

9 Pin

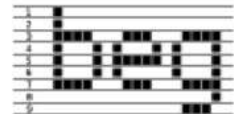
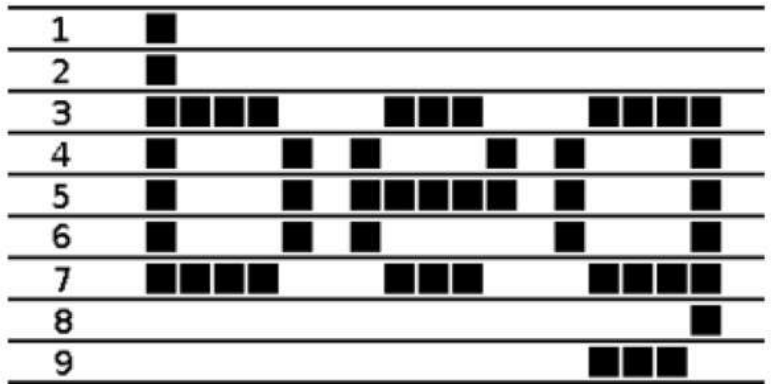
18 Pin

24 Pin

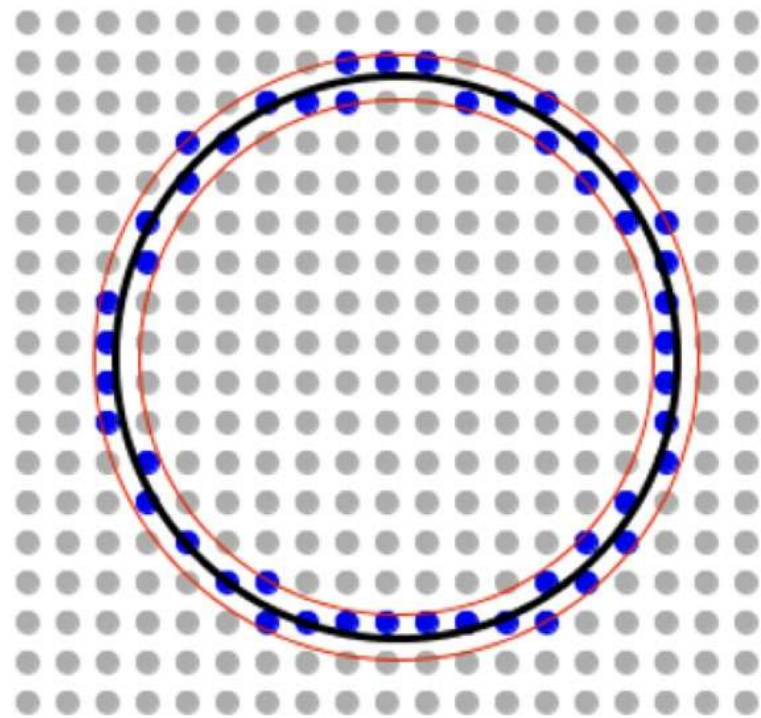
More Flexible, But Not a Cure-All



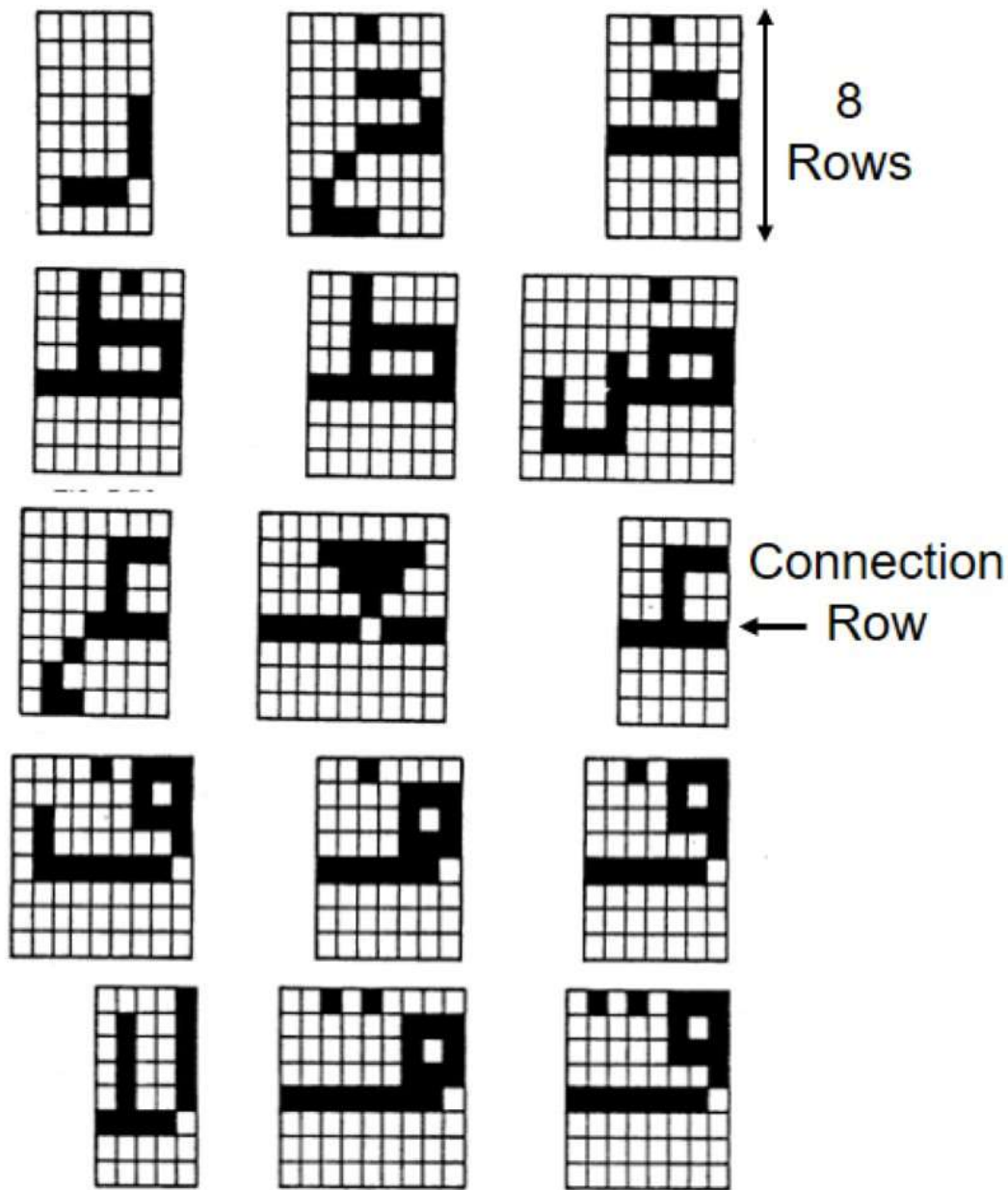
The same print mechanism usable for any language (software fonts) or for rendering graphic images



Pixel or digital geometry



Dot-Matrix Font Design



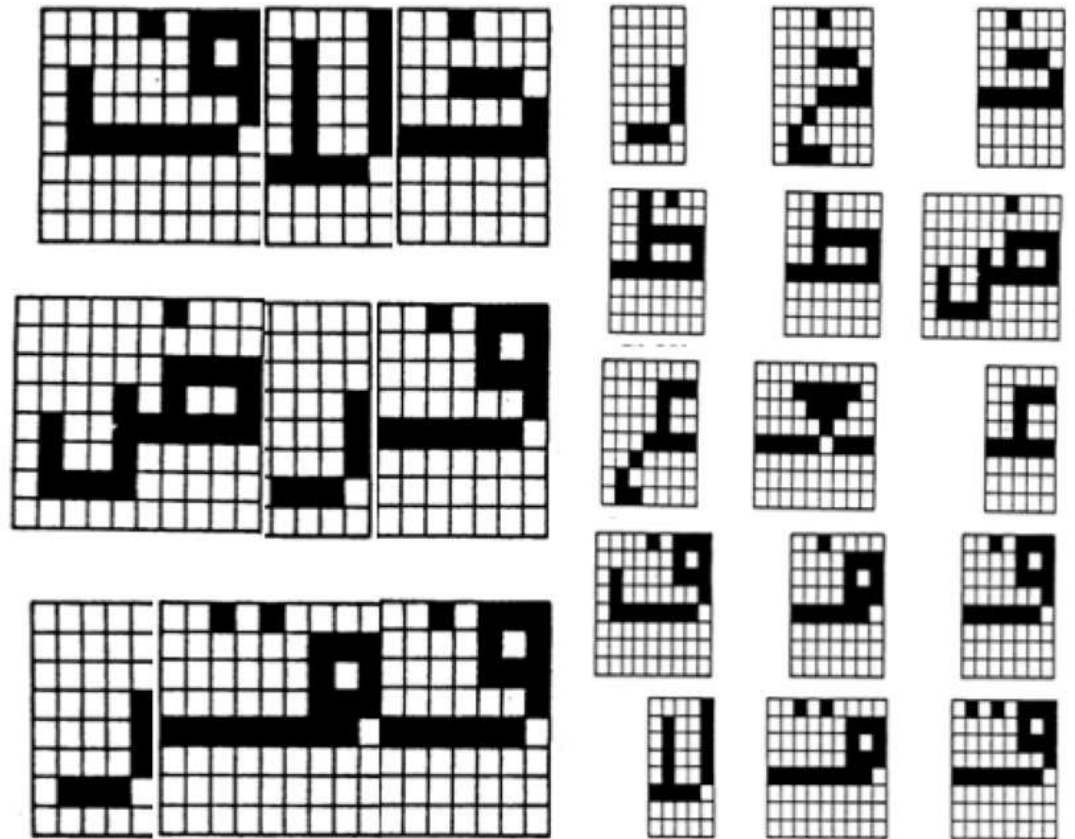
ABCDEFGHIJKLMNOP
 QRSTUVWXYZÀÁÂÃÄÅÆÇÈÉÊË
 bcdefghijklmnopqr
 stuvwxyzàáâãäåæçèéêë
 34567890(\$%&'()*+,-./:;<=>

abcdefghijklm
 nopqrstuvwxyz
 abcdefghijklm
 nopqrstuvwxyz

Persian Script Needs Larger Dot-Matrix

For a given dot-matrix size, legibility and pleasantness of the script is up to the font designer's skill and artistry

Characters must be viewed not in isolation, but in various combinations (with each other and with common punctuations)



Large Matrix + Small Dots = High Quality

Overlapped dots also help

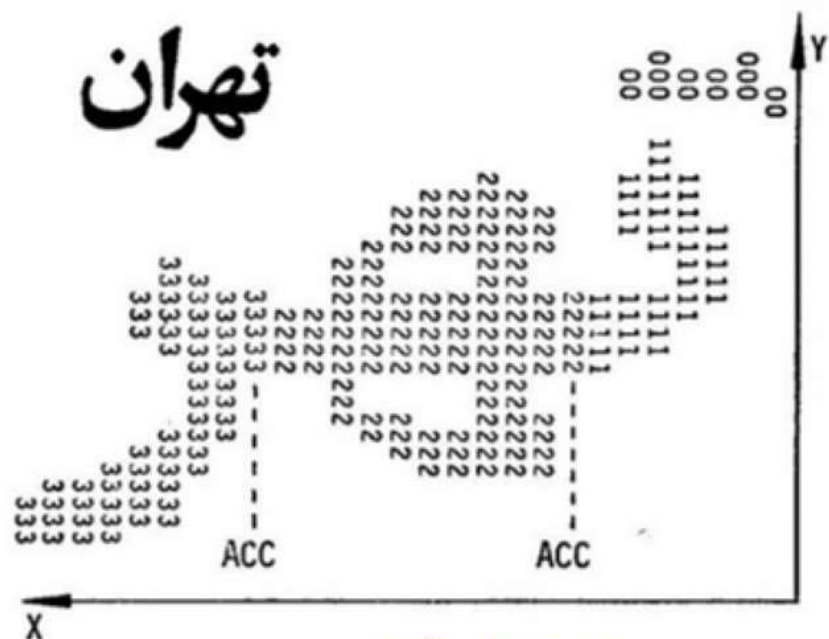
Variable widths is helpful



7 × 5 matrix,
disjoint dots



16 × 16 matrix,
overlapped dots



Improved Script with Dot-Matrix Printers

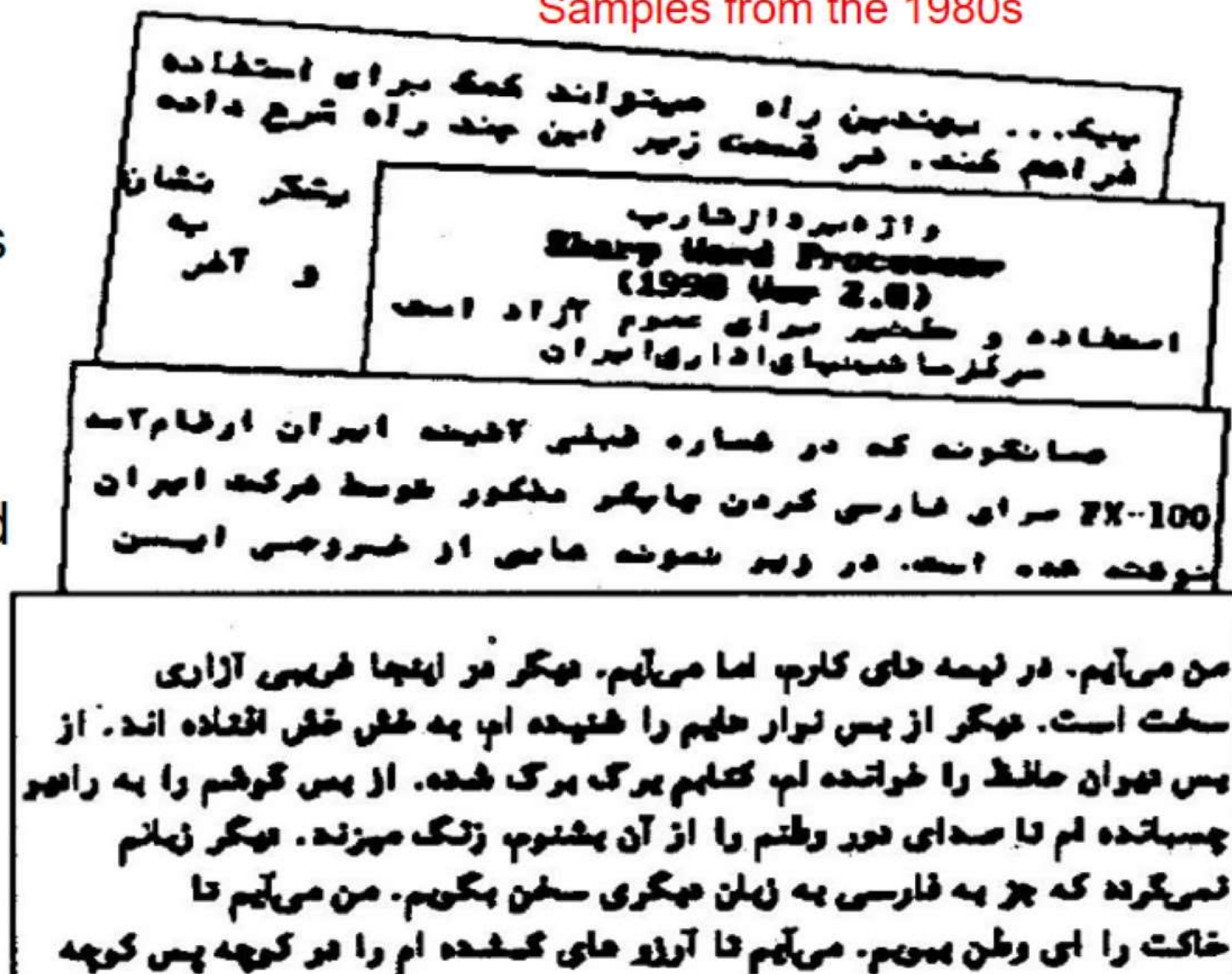
Variable widths
is helpful

Overlapped dots
also help

Highly flexible,
software-defined
font

Multiple fonts &
font weights,
with no added
hardware cost

Samples from the 1980s



Modern Persian Display Typefaces



Behrooz Parhami

1 hr · 🌐 ▼

کارِ پاکان را قیاس از خود مگیر
گر چه باشد در نوشتن شیر، شیر
آن یکی شیر است اندر بادیه
آن دگر شیر است اندر بادیه
آن یکی شیر است کادم می‌خورد
و آن دگر شیر است کادم می‌خورد

کارِ پاکان را قیاس از خود مگیر
آن یکی شیر است اندر بادیه
و آن دگر شیر است کادم می‌خورد

**Computer Persian
script samples**

کارِ پاکان را قیاس از خود مگیر
آن یکی شیر است اندر بادیه
آن یکی شیر است کادم می‌خورد

**Which one is
most legible?**

کارِ پاکان را قیاس از خود مگیر
آن یکی شیر است اندر بادیه
و آن دگر شیر است کادم می‌خورد

Samples from Facebook and other Internet sites



Behrooz Parhami

September 10, 2011 · 🌐 ▼

Selected verses from a poem by Hooshang Ebtehaj

عمری است تا از جان و دل ، ای جان و دل می خوانمت
کو نیز خواهان منی ، می دانمت ، می دانمت
ای خنده ی نیلوفر ی ، در گریه ام می آوری
بر گریه می خندی و من ، در گریه می خندانمت
ای زاده ی پندار من ، پوشیده از دیدار من
چو کودک نادانسته ، گپواره می جنبانمت
ای من کو بی من کیستی ، چون سایه بی من نیستی
همراه من می ایستی ، همپای خود می رانمت

دانش و فن آوری

حذف نرم افزارهای ایرانی از فروشگاه مجازی گوگل

📖 ویدئو: سلول های بنیادی

📖 محدودیت اینترنت: آزادی های آنلاین پس از انتخابات ۹۶

📖 چهل سال پس از وویجرها: دو نامه از یک تمدن

📖 محمدجواد آذری جهرمی و همزیستی شغلی یا فیلترینگ و

شنود

گوگل نیز پس از اهل
نرم افزارهای ایرانی را از
فروشگاه مجازی خود حذف
کرد. دلیل آن تبعیت از قانون
تحریم های ۲۲ ساله آمریکا
علیه ایران است.



Nov. 2017



Fifty Years of Poor Penmanship



Slide 38

Challenges in Persian Print Recognition

Technical Challenges

- Connectivity of letters
- Wide variations in letter widths/heights
- Context-dependent letter shapes
- Presence of 1-3 dots
- Small loops and other fine features
- Absence of short vowels (usually)



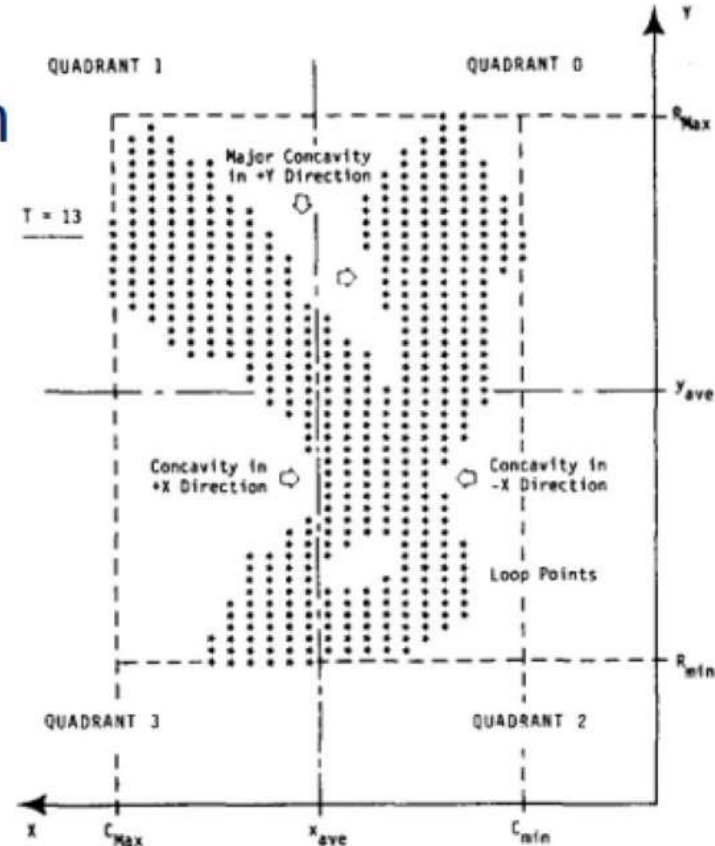
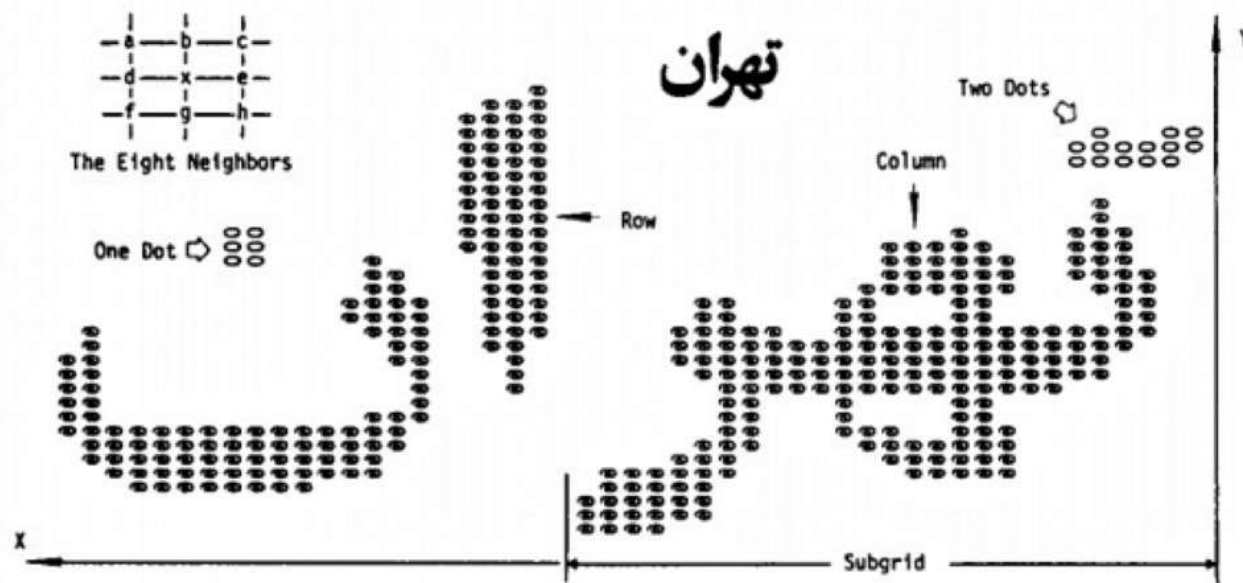
Five-Step Algorithm

- Decompose into lines
- Identify connection axis
- Decompose into letters
- Evaluate features
- Match extracted features

Persian Print Script Recognition

Examples of Features Used for Recognition

- Distribution of “weight” in the four quadrants
- Concavities (major, minor, in various directions)
- “Islands” (e.g., dots) and “lakes” (loops)
- Connectivity to the right or left symbol
- Row/column segment-length distribution



Lower Bound for Dot-Matrix Size

$$7 \times 5$$

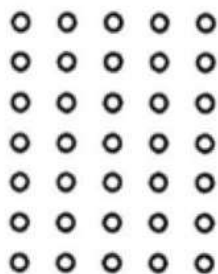
این جمله و ترکیبات زیر
نمایند: نوع خط حاصلند:
ابجد هوز طی کلیم ستغص
ترقت شخض طغخ آتیم شاک
چنگ حاذق دلائل عطش گیج
 $11293 + 75 \times 27 = 13048 \bar{R}$

Departures		13:52	
FLIGHT	DESTINATION	BOARDING TIME	GATE STATUS
NZ119	SYDNEY	12:20	07 DEPARTED
FJ410	NADI	13:10	01 GATE CLOSED
QF190	SYDNEY	13:10	04 GATE CLOSED
NZ175	PERTH	13:25	10 FINAL CALL
MH130	KUALA LUMPUR	13:30	03 FINAL CALL
CX108	HONG KONG	13:40	08 BOARDING
QF126	BRISBANE	13:50	07 BOARDING IN 5 MINS
NZ725	MELBOURNE	14:50	BOARDING IN 60 MIN
NZ739	BRISBANE	14:50	BOARDING IN 60 MIN
TG990	BANGKOK	14:50	BOARDING IN 60 MIN
NZ105	SYDNEY	16:05	PLEASE WAIT
	MELBOURNE	15:45	PLEASE WAIT

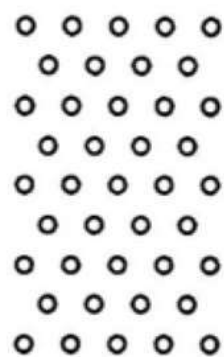
$$7 \times 9/2$$

این جمله و ترکیبات زیر
نمایند: نوع خط حاصلند:
ابجد هوز طی کلیم ستغص
ترقت شخض طغخ آتیم شاک
چنگ حاذق دلائل عطش گیج
 $11293 + 75 \times 27 = 13048 \bar{R}$

$$7 \times 5$$



$$9 \times 9/2$$



ملاحظه کاری، طبخ غذا، خطاطی، گل آلاله

ملاحظه کاری، طبخ غذا، خطاطی، گل آلاله

ملاحظه کاری، طبخ غذا، خطاطی، گل آلاله

ملاحظه کاری، طبخ غذا، خطاطی، گل آلاله

Tall letters with 7×5 , $7 \times 9/2$,
 $9 \times 9/2$, and $9 \times w$ matrices

$$9 \times 9/2$$

این جمله و ترکیبات زیر
نمایند: نوع خط حاصلند:
ابجد هوز طی کلیم ستغص
ترقت شخض طغخ آتیم شاک
چنگ حاذق دلائل عطش گیج
 $11293 + 75 \times 27 = 13048 \bar{R}$

Scalability of Persian Fonts and Script

The best response to fools is silence

Arial 28

جواب ابلهان خاموشی است

Arial 28

جواب ابلهان خاموشی است

32

جواب ابلهان خاموشی است

36

جواب ابلهان خاموشی است

40

The best response to fools is silence

Arial 14

جواب ابلهان خاموشی است

Arial 14

جواب ابلهان خاموشی است

16

جواب ابلهان خاموشی است

18

جواب ابلهان خاموشی است

20

The best response to fools is silence

Arial 14

جواب ابلهان خاموشی است

Arial 9

جواب ابلهان خاموشی است

10

جواب ابلهان خاموشی است

11

جواب ابلهان خاموشی است

12

**Font-size
adjustment**

The best response to fools is silence

Arial 28

جواب ابلهان خاموشی است

Arial 28

جواب ابلهان خاموشی است

32

جواب ابلهان خاموشی است

36

جواب ابلهان خاموشی است

40

The best response to fools is silence

Arial 28

جواب ابلهان خاموشی است

Arial 28

جواب ابلهان خاموشی است

32

جواب ابلهان خاموشی است

36

جواب ابلهان خاموشی است

40

**Image
resizing**

Microsoft Word Persian Fonts

All fonts 32, converted to JPG from MS Word doc

خط

خط

خط

خط

Same

Arial خط فارسی مایکروسافت وُرد

Cambria خط فارسی مایکروسافت وُرد

Helvetica خط فارسی مایکروسافت وُرد

Times New Roman خط فارسی مایکروسافت وُرد

Calibri خط فارسی مایکروسافت وُرد

Courier خط فارسی مایکروسافت وُرد

Dubai خط فارسی مایکروسافت وُرد

Dubai Light خط فارسی مایکروسافت وُرد

Tahoma خط فارسی مایکروسافت وُرد

Traditional Arabic خط فارسی مایکروسافت وُرد

Where We Are, Where We Go from Here

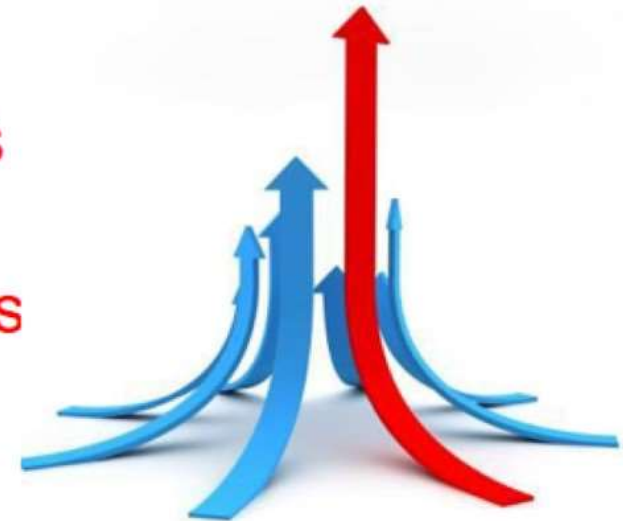
Technology to the rescue

- Improved resolution/contrast in displays
- HD (1920 x 1080) and UHD display units
- Obsolescence of line-segment displays
- High-res printers (600 dpi or more)
- Accumulated know-how about font design



Path to further improvements

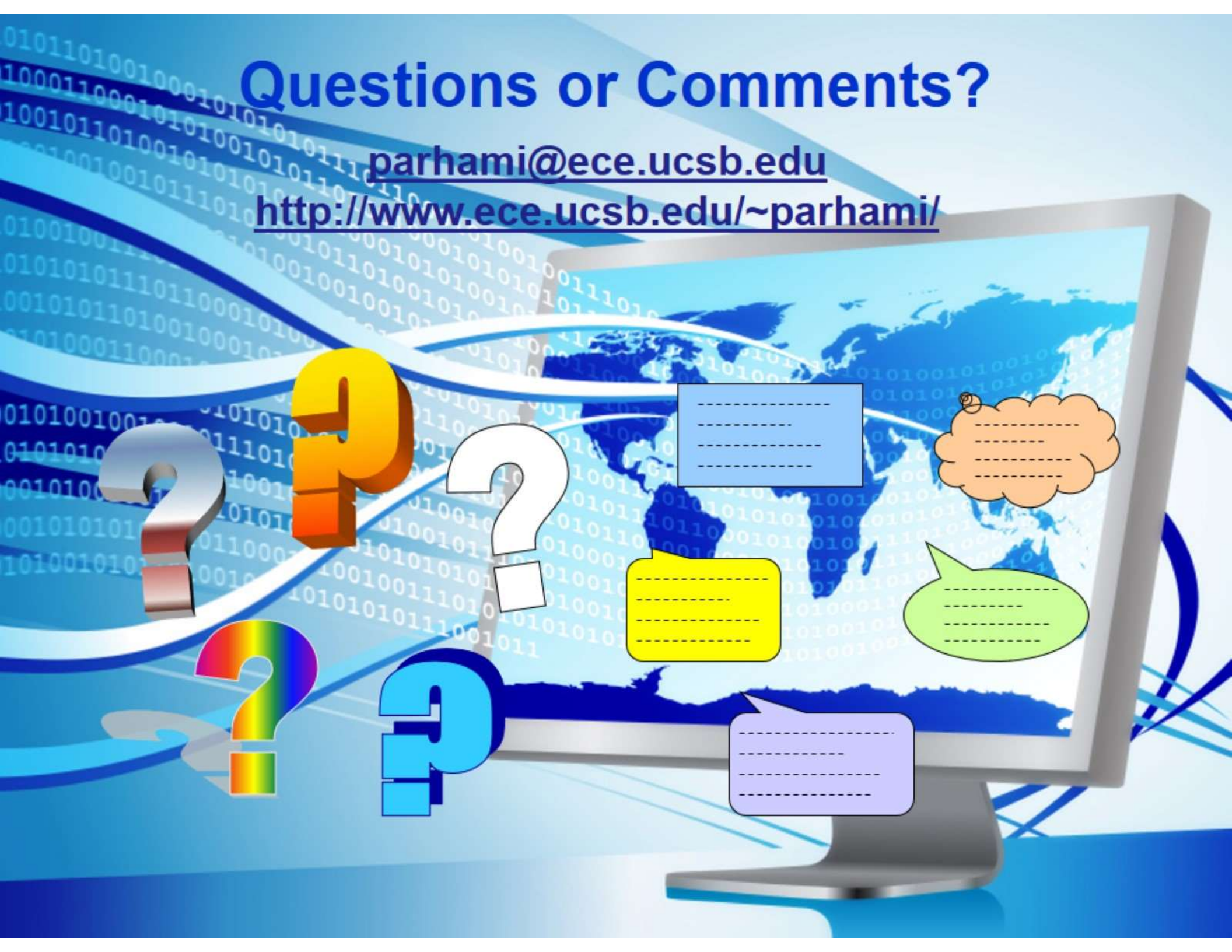
- Trade-offs between legibility and aesthetics
- Learning from large data sets
- Crowd-sourcing the evaluation of aesthetics
- Formatting, especially in bilingual output
- Expect more to get more!



Questions or Comments?

parhami@ece.ucsb.edu

<http://www.ece.ucsb.edu/~parhami/>



Back-up Slides

2010-17

2000s

1990s

1980s

1970s

1967-69



Behrooz Parhami

University of California, Santa Barbara

مروری بر تاریخچه‌ی تولید خط فارسی توسط کامپیوتر



بهر روز پرهامی

دانشگاه کالیفرنیا در سنتا باربارا

Persian Handwriting

Examples of beautiful handwriting

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
 گدازنده من و منان پروردگارم و فدای خاکشده شد، این قلم است که در دست تو بود و در دست من و در دست او
 بر تخت آرد و نشیند - گویند که این قلم را در دست من و در دست تو و در دست او نشیند و در دست او نشیند
 مردم با او نشیند و این قلم را در دست او نشیند و در دست او نشیند و در دست او نشیند
 و در دست او نشیند و در دست او نشیند و در دست او نشیند و در دست او نشیند
 مستخدم حرف می آید و در دست او نشیند و در دست او نشیند و در دست او نشیند
 گشته و در دست او نشیند و در دست او نشیند و در دست او نشیند و در دست او نشیند
 بلکه در دست او نشیند و در دست او نشیند و در دست او نشیند و در دست او نشیند
 باشد و در دست او نشیند و در دست او نشیند و در دست او نشیند و در دست او نشیند
 با تو و در دست او نشیند و در دست او نشیند و در دست او نشیند و در دست او نشیند
 آنچه در دست او نشیند و در دست او نشیند و در دست او نشیند و در دست او نشیند
 مفاصلی که ۱۳۶۱

عشق
 بهترین هدیه است
 به هر کس که دوست دارد
 ۱۳۶۱

A schoolchild's handwriting



Ali-Akbar Dehkhoda's handwriting

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
 این قلم را که در دست من و در دست تو و در دست او نشیند
 و در دست او نشیند و در دست او نشیند و در دست او نشیند
 و در دست او نشیند و در دست او نشیند و در دست او نشیند
 و در دست او نشیند و در دست او نشیند و در دست او نشیند