



Developing Multilingual OCR and Handwriting Recognition at Google

Observations and Reflections

Ashok Popat

Research Scientist, Google Inc.

IAPR Summer School, Jaipur: Jan 23 2017



Joint work with

Jon Baccash

Marcos Calvo

Victor Cărbune

[Thomas Deselaers](#)

Karel Driesen

Sandro Feuz

[Yasuhisa Fujii](#)

Philippe Gervais

Pedro Gonnet

Patrick Hurst

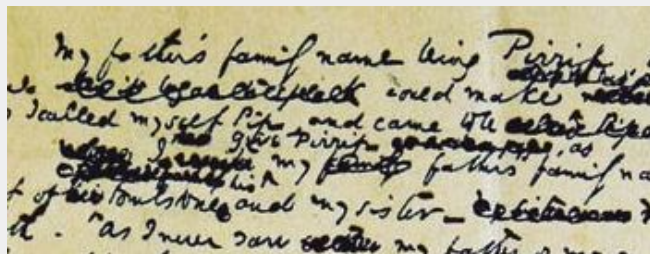
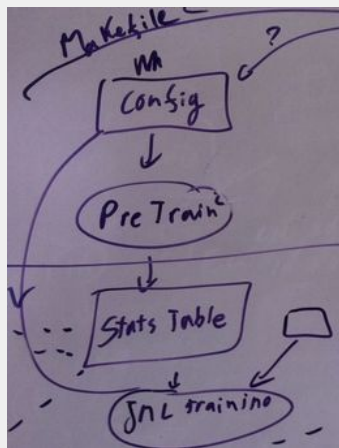
Henry Rowley

Li-Lun Wang

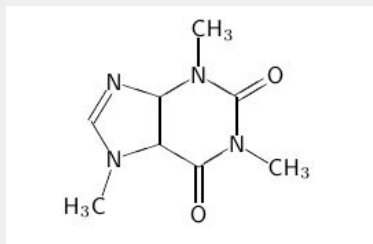
Optical Character Recognition

*Acona, gwana acona n^3 : Gwana ma kodine dong orweny woko. Acona, gin macon - nying dano ma laco - twero bedo

ber.
Acwicwini, acwicwino n^2 : Lum ma lotino gijwatte nyo gipwodde kwede; lotino



$$s[n] \stackrel{\text{def}}{=} T \int_{-1/T}^{1/T} S_{1/T}(f) \cdot e^{i2\pi f n T} df = T \underbrace{\int_{-\infty}^{\infty} S(f) \cdot e^{i2\pi f n T} df}_{\stackrel{\text{def}}{=} s(nT)}$$

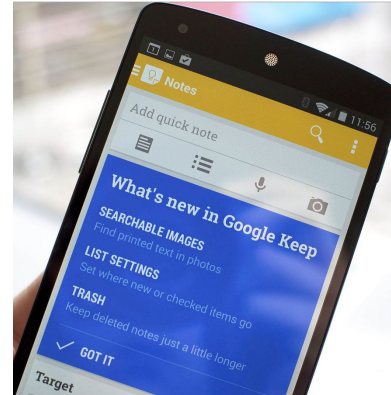




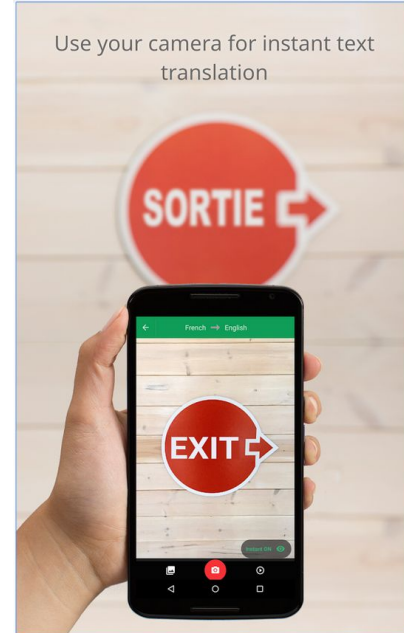
OCR in Google Products



Google Keep - notes and lists

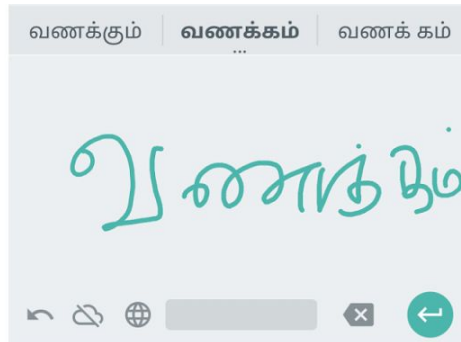
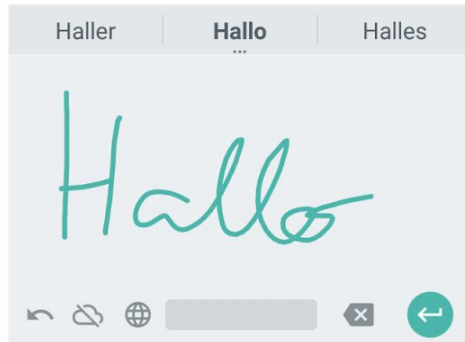
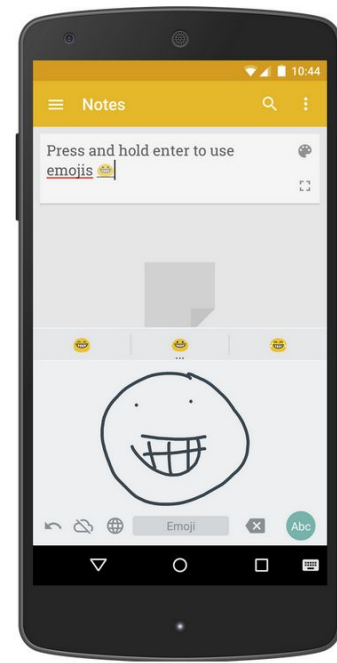
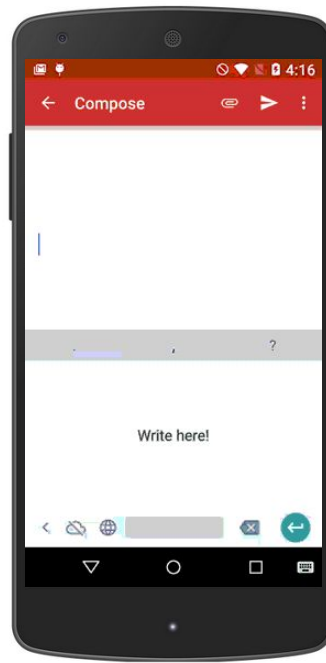


Google Translate



Google Handwriting Input

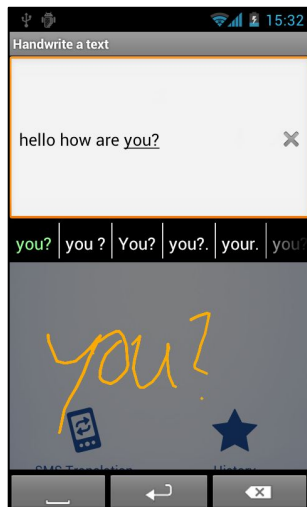
on-device recognition
> 80 languages + emoji



Google Translate for Android

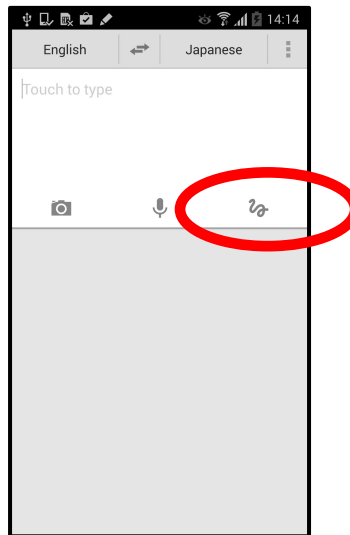
Translate 2.3

enabled by default only for CJ



Translate 2.4+

enabled for all supported lang.



Handwrite for Mobile Search

write your search right on the
Google homepage

available on Google.com from
smartphone or tablet

can be activated or disabled in
mobile search settings

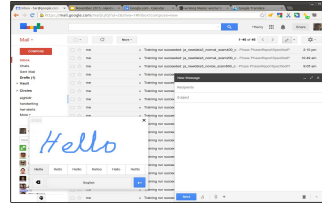


Other Applications

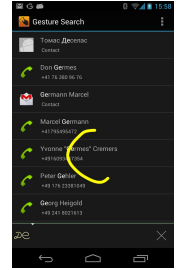
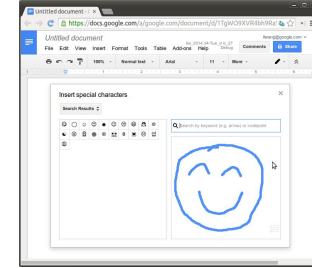
Other input methods for Android



Input tools



... and more



Outline

- Multilingual OCR and On-line handwriting systems
- Research at Google
- Personal observations, reflections



Part 1a: A multilingual OCR system

Examples from Google Books

Multiple scripts / languages on a page:

(601, 602.) CIVITAS. LIB. X. 181

χρωμάτων δὲ καὶ σχημάτων θεωροῦσι; Πάνυ μὲν οὖν. Οὕτω δὲ, οἶμαι, καὶ τὸν ποιητικὸν φήσομεν χρώματα ἅττα ἐκάστων τῶν τεχνῶν τοῖς ὀνόμασι καὶ ῥήμασιν ἐπιχρωματίζειν, αὐτὸν οὐκ ἐπαίοντα ἀλλ' ἢ μιμῆσθαι, 5 ὥστε ἐτέροις τοιοῦτοις ἐκ τῶν λόγων θεωροῦσι δοκεῖν, εἴαν τε περὶ σκυτοτομίας τις λέγῃ ἐν μέτρῳ καὶ ῥυθμῷ καὶ ἁρμονίᾳ, πάνυ εὖ δοκεῖν λέγεσθαι, εἴαν τε περὶ στρατηγίας εἴαν τε περὶ ἄλλου ὁποῦν· οὕτω φύσει αὐτὰ ταῦτα 10 μεγάλῃ τινὰ κλησίν ἔχειν. ἐπεὶ γυμνοθέντα γε τῶν τῆς μουσικῆς χρωμάτων τὰ τῶν ποιητῶν, αὐτὰ ἐφ' αὐτῶν λεγόμενα, οἶμά σε εἰδέναι οἷα φαίνεται. τεθέασαι γάρ που. Ἔγωγ', ἔφη. Οὐκ οὖν, ἦν δ' ἐγὼ, εἴκει τοῖς τῶν ὠραίων 15 προσώποις, καλῶν δὲ μὴ, οἷα γίγνεται ἰδεῖν, ὅταν αὐτὰ τὸ ἄνθος προλίπη; Παντάπασιν, ἦ δ' ὅς. Ἴθι δὲ, τόδε ἄθρει· ὁ τοῦ εἰδῶλου ποιητῆς, ὁ μιμητῆς, φάμεν, τοῦ μὲν ὄντος οὐδὲν ἐπαίει, τοῦ δὲ φαινομένου. οὐχ οὕτω; Ναί. Μὴ τοίνυν ἡμίσεως αὐτὸ καταλίπομεν 20 ῥηθὲν, ἀλλ' ἱκανῶς ἴδωμεν. Λέγε, ἔφη. Ζωγράφος, (socr.) Haud secus, opinor, poetam dicemus colores quosdam artium singularum nominibus verbisque exprimere, cum ipse nihil sciat nisi imitari, adeo ut aliis similibus secundum verba spectantibus bene dicere videatur, sive de sutrina dicat versibus et numero et harmonia, sive de re militari, seu de quibuslibet aliis: tantam naturam in his ipsis inesse delectationem. nam musicæ coloribus nudata 25 poemata per se sola spectata scire te arbitrator qualia videantur. vidisti enim, nisi fallor. [cl.] Vidi, inquit. [socr.] Nonne perinde se habent, inquam, ac facies eorum, qui in flore ætatis constituti neque pulchri sunt, quales visu fiunt, quando flos eis decedit? [cl.] Plane, inquit. [socr.] Nunc, age, hoc considera: imitator, hoc est, simulacri auctor, veri cognitionem nullam habet, sed visi. nonne? [cl.] Ita est. [socr.] Ne igitur semiperfectum id relinquamus, sed plene perspiciamus. [cl.] Dicam, inquit. [socr.] Pictor, dicimus, habenas pinget et frenum? [cl.] Ita. [socr.] Faciet autem coriarius et faber ærarius? [cl.] Faciet. [socr.] Num igitur


Examples from Google Books (cont.)

Per-word script and language variation:

Dieser Familie nähern sich die Slawonischen Worte *лана*, *лопастъ* und vorzüglich *лопата* (*lapa*, *lopast* und *lopata*, Pfote, Flügel einer Mütze, Schaufel), mit welcher ein Blatt durch seine Fläche eine grosse Aehnlichkeit hat. Daher nennen auch wir die grossen Blätter einiger Gewächse *лапушникъ* (*lapuschnik*, Klette). Wie sehr sich aber auch die letztern Worte

Examples from Google Books (cont.)

§. 6. Vesezeichen:

1. virâma, „Ruhezeichen“ , ein unten rechts an den Konsonanten angefügter Strich, nimmt dem Konsonanten den a-Laut, mit welchem er nach § 1 und § 4 an und für sich zu sprechen ist. Der Virâma findet sich namentlich am Ende eines Wortes und manchmal anstatt einer Ligatur; z. B.

यत् yat, आसीत् âsit, आगमद् देवी âgamad dêvi;
क्त्र ktra kann auch geschrieben werden क्त्वर.

2. avagraha, „Apostroph“ ऽ, steht am Anfange eines Wortes zur Bezeichnung eines vorn abgefallenen a: z. B. दुःखितो
ऽभवत् duḥkhitô 'bhavat.

Some of the 26 scripts of interest

據以上對環山青年的價值青年對目前工作感到高度社會的調適卻不是有利的了適應正在急遽變遷的現狀環境而不為環境所控制與高度的依賴性格則明顯的經濟也不是恆久不變的因素

terpretácie umeleckého dieťa. Čenené predovšetkým laickým A. Konceptiu umeleckého obsahu, ktorý jestvuje etňuje a vyjadruje (Dvořák) ou diela je práve toto sprisahujú. Spomenutými furritomňovanie a konkretizovaný obsah zahafujú. Akcupňuje historickou povahou ozumiteľné. A ideová pods

والفضاء . ونلاحظ أنه ف
الآن تقاس بالسنوات المفر
ث حقب حضارية بدلاً م
الواحد لا يشكل إلا جز
الوقت نلاحظ ان كل حق
ناحى الحياة الإنسانية تزا

սկունն ես առաջ էլ եմ հ
գործի մեջ, որ վերաբը
ված բերդապահ գործի հ
սզի զինվորները ձեր
սրզանք տածեն դեպի ա
հսակ թույլ վերաբերմուն
ները պարտավորվել են
ենց այդ պարտավորութ
պարտավորութունների հե
ղնում էր այն պատկերաց

अटलुं स्वीकारवुं रलुं के 7मी सदी
॥भोअे वीश्वकक्षाना वीद्वानो ज्ञान
कता. युरोपमां जे जगृती आवेली
, दकतरी वीधा, अवकाशज्ञान,
वशाशास्त्रमां करेली प्रगतीने आत्म
र ठेर स्थापत्यना उत्कृष्ट नमुना
महाल, गोलगुंजम, यार भीन
डेरी आयोजन, स्थापत्य, नागरीकर

發展をはかることにある。
は、拠点開発方式により総合開発
のため、全国の区域を政策対象の
、開発地域の3地域に区分し、過
度集中を防止し、再開発をはかる
を行なうものとし、整備地域につ
、主として計画的に工業分散を誘
発地域については、大規模地方開
、積極的に開発を促進するための
開発効果を通じて産業および都
全般にわたる有効適切な利用をは

Starting point: Markov-model-based approaches

- Document image decoding [Kopec and Chou, 1994]
 - Explicit model of typesetting process: seek to invert
 - Influenced by speech recognition methods
 - Extremely high accuracy when models match the data

- BBN Byblos system [Schwartz et al., 1996]
 - Treat text line like a speech waveform
 - Built on existing speech recognition system
 - First successful Arabic OCR

Generalization of the noisy channel model

- Speech approach

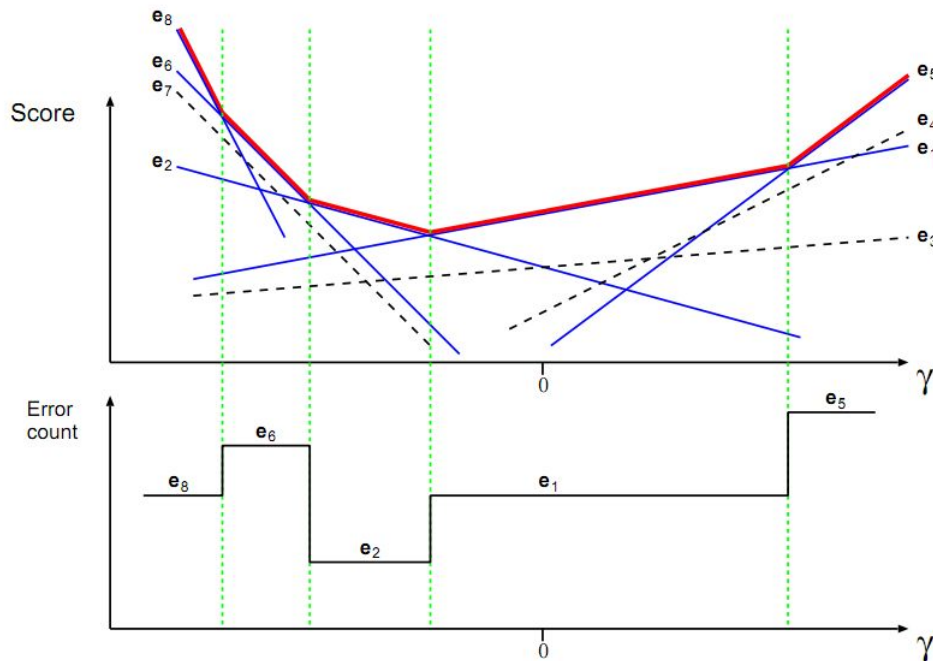
$$\begin{aligned} m' &= \arg \max_m P(m)^\alpha P(x|m) \\ &= \arg \max_m \alpha \log P(m) + \log P(x|m) \end{aligned}$$

- Generalize to multiple feature functions

$$m' = \arg \max_m \sum_i \lambda_i h_i(x, m)$$

- Learn $\{\lambda\}$ via minimum error-rate training [Macherey et al. '08, Och '03]

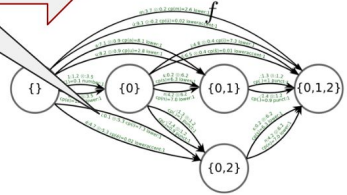
Minimum Error Rate Training



Label "i"
 Feature functions values:
 0.1 - character score
 0.9 - language model score
 2.3 - relative size to neighbors
 0.2 - cut score
 Label "e"
 [...]

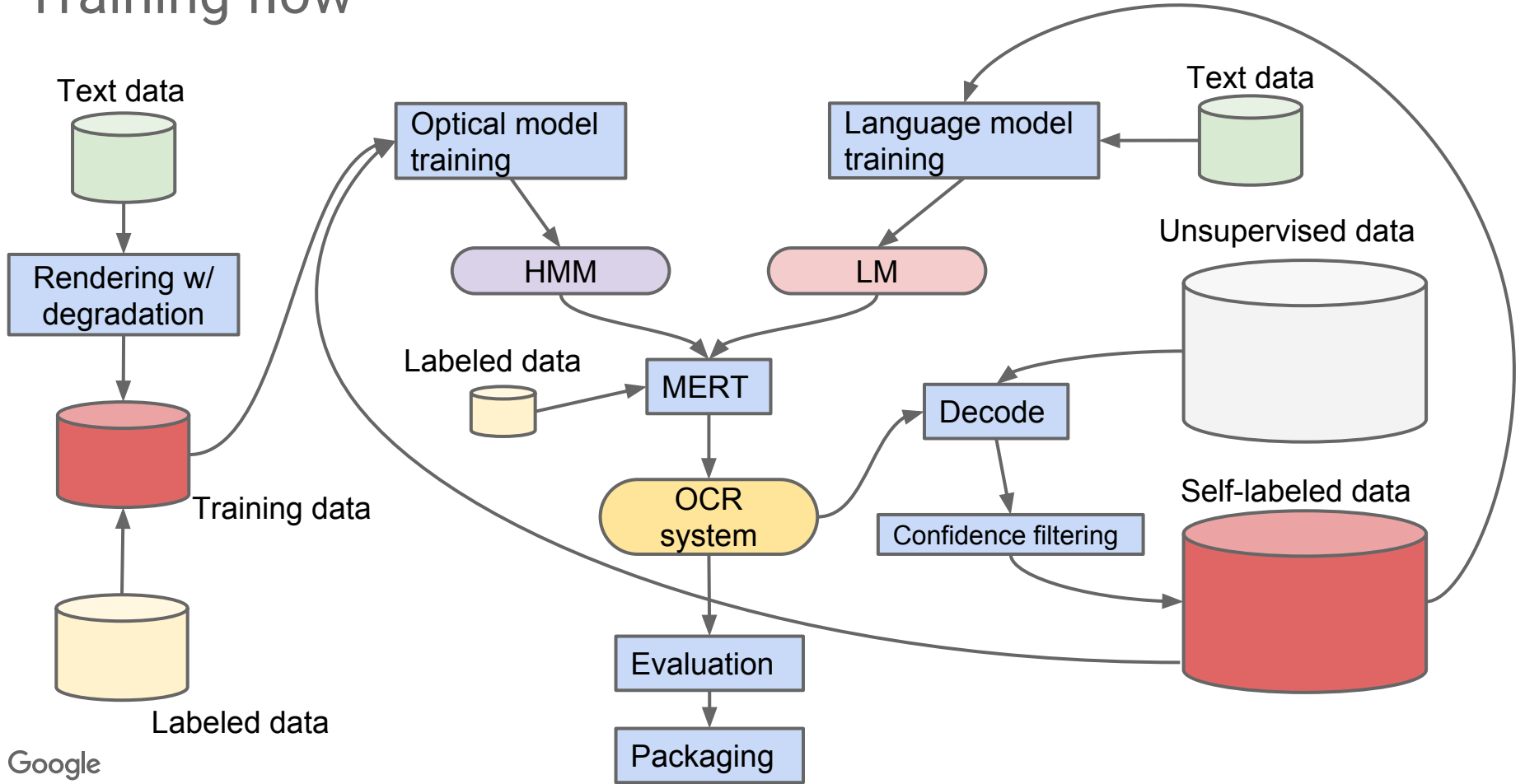
determine edge score as weighted sum

$$c = \sum_f \gamma_f c_f$$



Macherey, Och, Thayer, Uszkoreit: Lattice-based Minimum Error Rate Training for Statistical Machine Translation. EMNLP 2008.

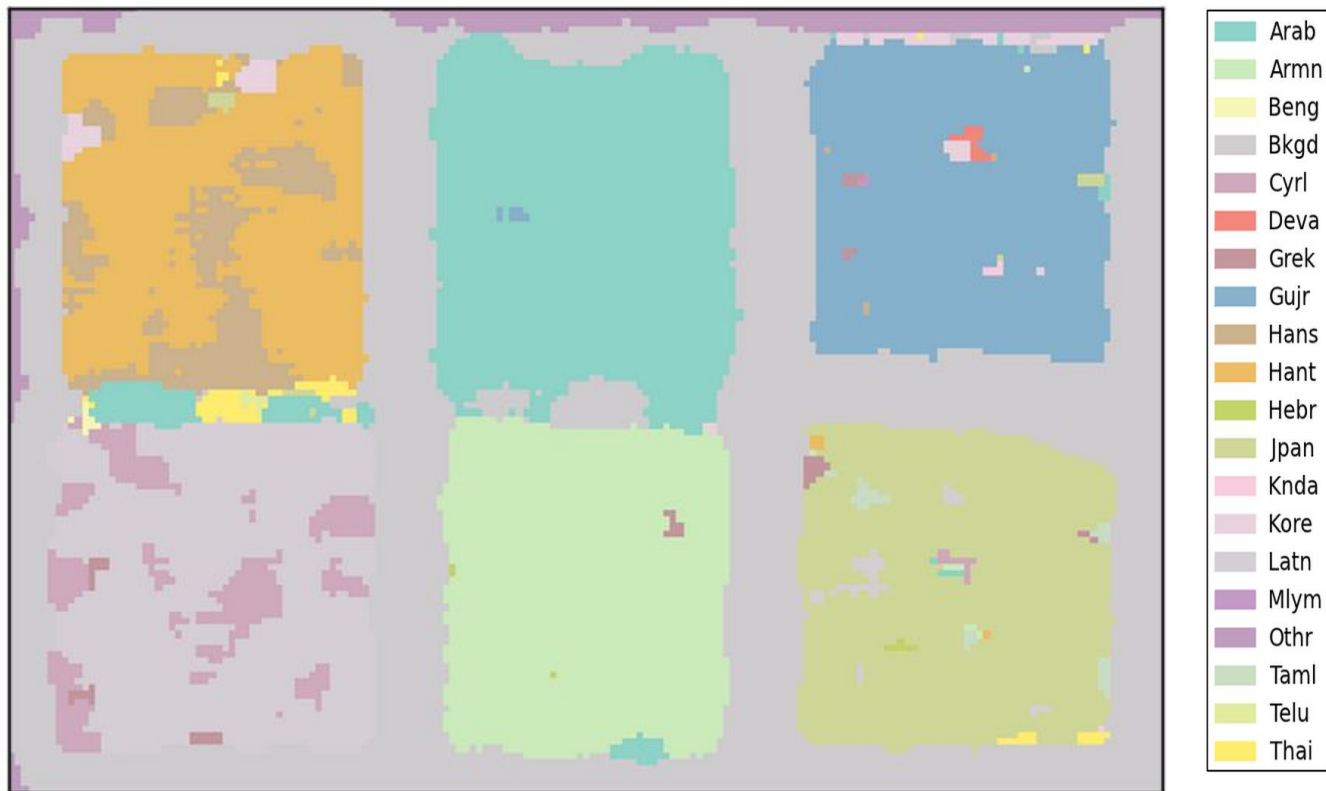
Training flow



Technical evolution

- Optical model
 - GMM -> DNN
 - DNN -> LSTM
 - Sequential discriminative training of DNN/LSTM
- Language model
 - N-gram -> RNN-LM
- Decoding
 - Pruning algorithms designed for OCR
 - Automatic decoding parameter optimization
 - Fujii et al., ICDAR'15

Script ID (Li et al., 2015)



Regions not covered

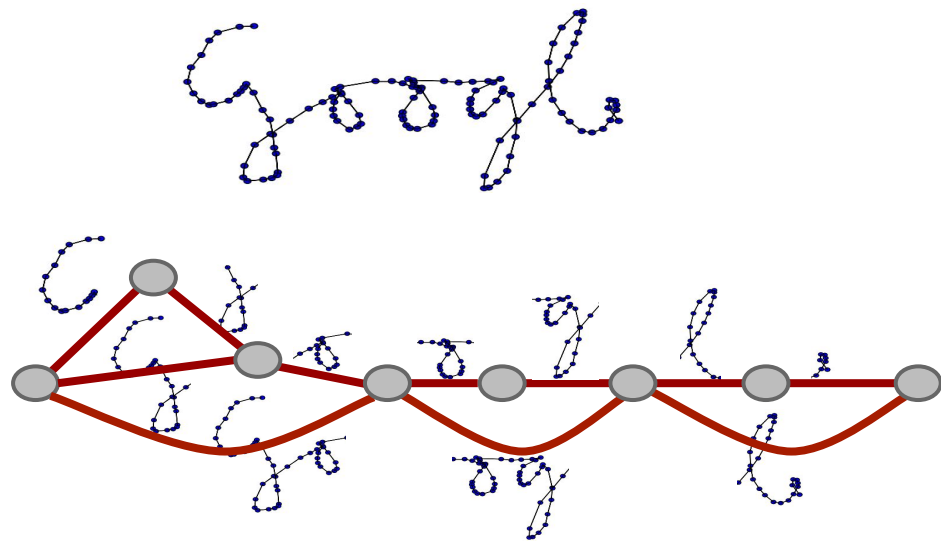




Part 1b: A multilingual handwriting recognition system



Segment and Decode

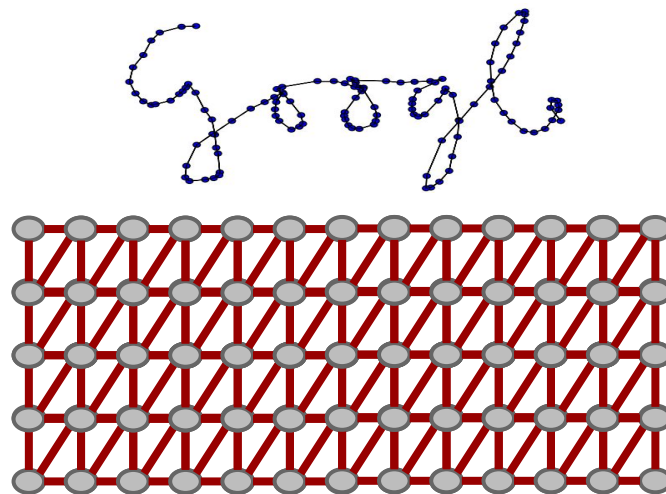


Apple Newton [Yaeger 1996]

Microsoft Tablet PC / Vista [Pittman 2007]

Google

Hidden Markov Models



neural network variants:

Recurrent, Time-Delay,

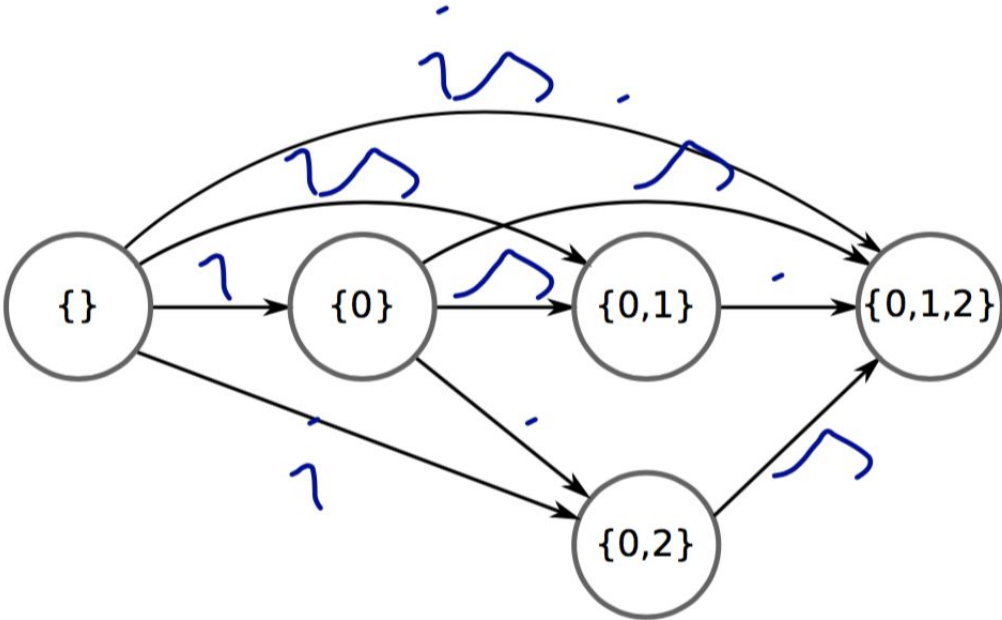
Long Short-term Memory

[Jaeger 2001], [Graves 2009], ...

Segment and Decode 1: Creating a segmentation lattice

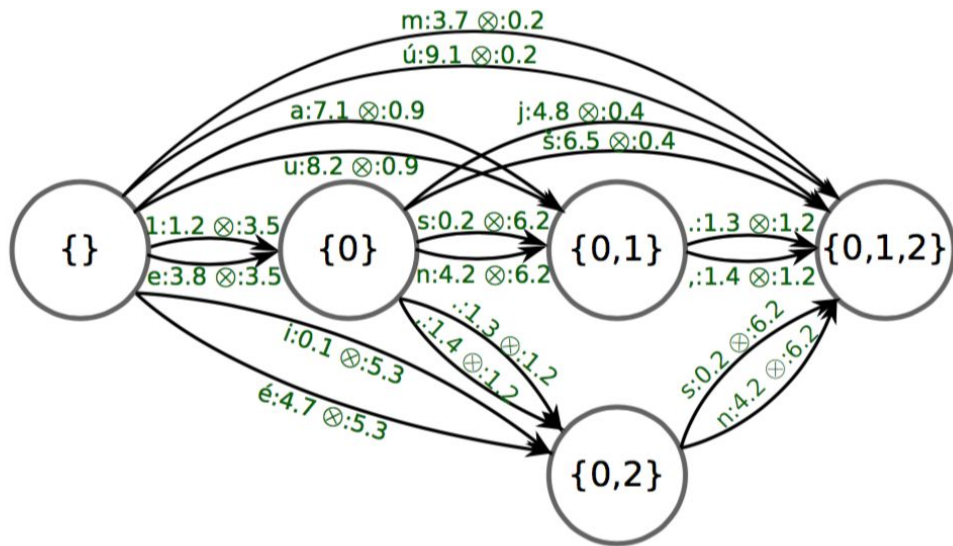
is

2
0 1

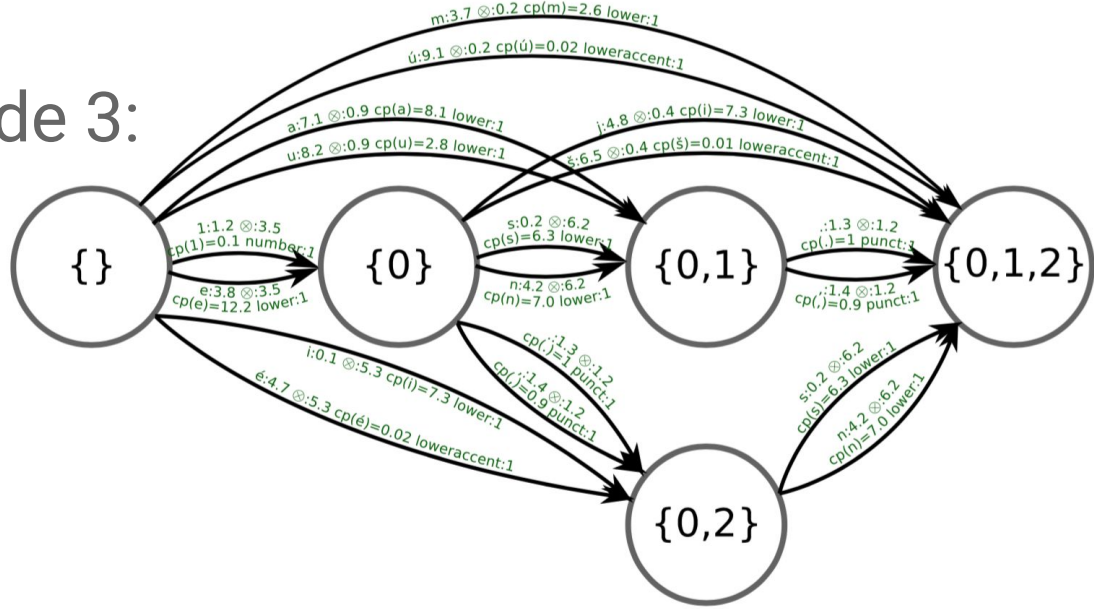


Segment and Decode 2: recognizing character hypotheses

2, 0		l:1.2 e:3.8 ⊗:3.5
1		s:0.2 n:4.2 ⊗:6.2
2		.:1.3 ,:1.4 ⊗:1.2
0,1		a:7.1 u:8.2 ⊗:0.9
0,2		i:0.1 é:4.7 ⊗:5.3
1,2		j:4.8 š:6.5 ⊗:0.4
0,1,2		m:3.7 ú:9.1 ⊗:0.2



Segment and Decode 3: Decoding



Beam 0 steps:				
edge	lm	total		
label	cost	cost	cost	path
	0.0	0.0	0.0	{}

Beam 1 steps:				
edge	lm	total		
label	cost	cost	cost	path
1	-1.2	9.7	8.5	{}-0
i	3.1	7.3	10.4	{}-02
é	0.4	13.5	13.9	{}-02
m	7.1	13.4	20.5	{}-012
u	11.1	9.9	21.0	{}-01
e	13.5	7.9	21.4	{}-0
ú	9.9	12.0	21.9	{}-012
a	15.3	7.0	22.3	{}-01

Beam 2 steps:				
edge	lm	total		
label	cost	cost	cost	path
is	1.3	8.3	19.9	{}-02-012
1.	2.1	9.4	20.0	{}-0-02
m			20.5	{}-012
1,	2.1	10.5	21.1	{}-0-012
1s	1.3	13.5	23.3	{}-0-01
in	6.0	7.9	24.2	{}-02-012
1n	6.0	14.0	28.5	{}-0-01
és	1.3	17.7	32.9	{}-02-012
1j	12.7	14.0	35.2	{}-0-02
1š	7.1	24.0	39.6	{}-0-02
én	6.0	19.9	39.8	{}-02-012

Beam 3 steps:				
edge	lm	total		
label	cost	cost	cost	path
is			19.9	{}-02-012
m			20.5	{}-012
1.s	1.3	36.3	37.6	{}-0-02-012
1,s	1.3	41.2	42.5	{}-0-02-012
1.n	6.0	39.2	45.2	{}-0-02-012
1,n	6.0	40.1	46.1	{}-0-02-012

Feature Function Weights

Label "i"

Feature functions values:

0.1 – character score

0.9 – language model score

2.3 – relative size to neighbors

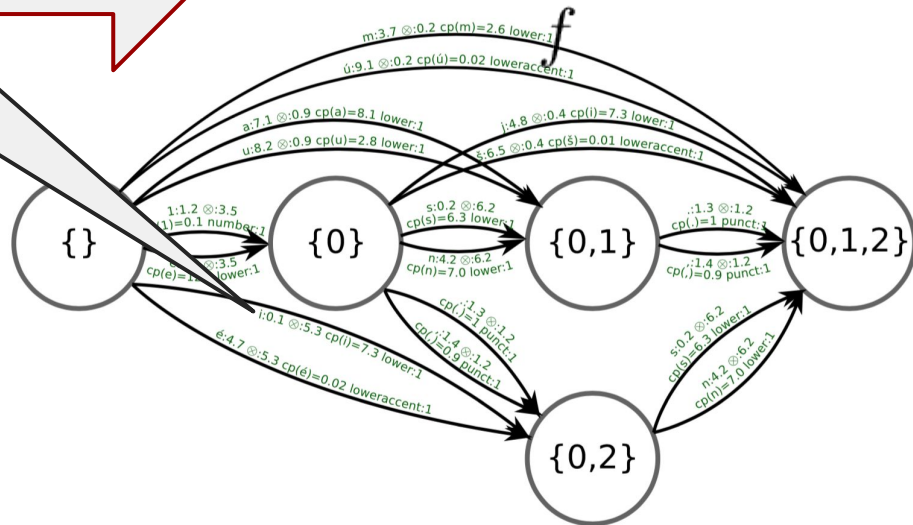
0.2 – cut score

Label "é"

[...]

determine edge
score as
weighted sum

$$c = \sum_f \gamma_f c_f$$



Features: Per character hypothesis

- Histograms of point features (3210 dimensional)
- Bitmap features: 3x8x8 pixels (192 dimensional)
- Simple statistics (384 dimensional)
- Water reservoir features (64 dimensional)
- Stroke direction (180 dimensional)
- Quantized stroke direction maps (512 dimensional)

More feature functions

- string length
- character prior
- segmenter cut features
- relative size



Part 2: Research at Google



Google's Hybrid Approach to Research

Spector, Norvig, Petrov '12 *Comm. of the ACM*

- Pattern 2: Small research team builds a system that gets deployed.

“This pattern applies best when continuing research can further improve and extend the resulting products.”

Enablers

- Single code base, wide range of library functions
- Infrastructure
- Expertise and skills of other teams
- Data

Enablers (cultural)

- Transparency and cooperation
- Peer review
- Respect and psychological safety
- Team- and personal-level pace and execution
- Data-centrism

Software engineering

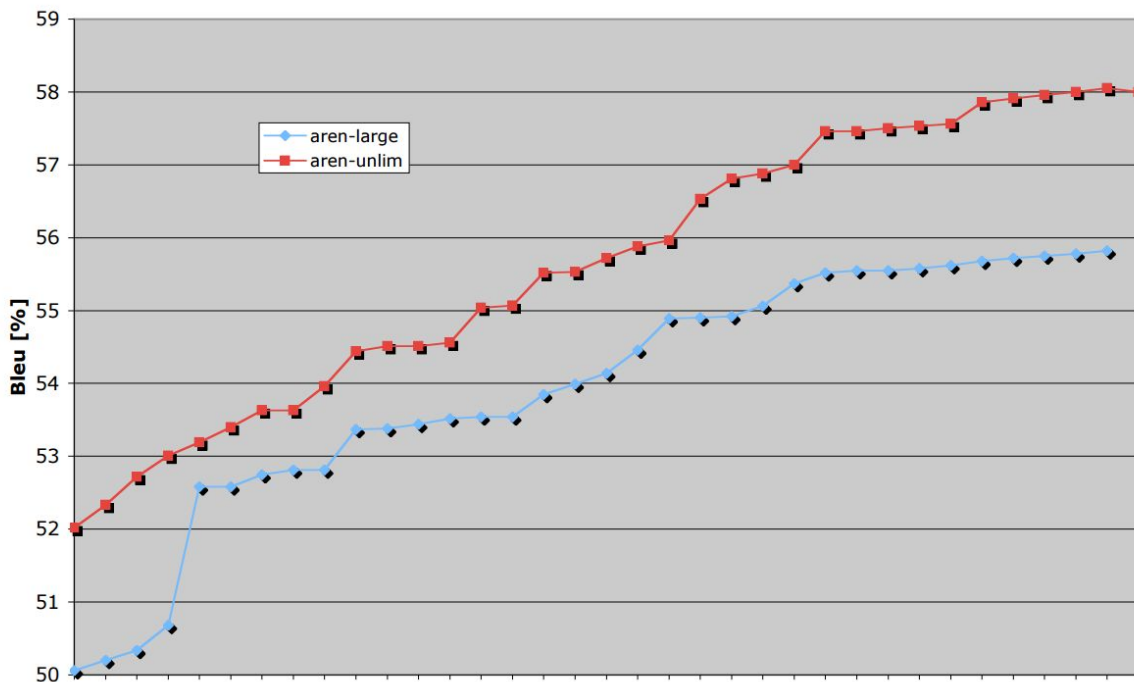
- Respected and valued
- If it's not checked in, it doesn't exist
- Toy prototypes versus production-quality code
- A day in the life: 80/20



Part 3: Observations and Reflections



Improvement over time (May - July): Arabic



Rapid real progress

- Multiple contributors, one system
- Industry folks at NIST'06 meeting were startled
- Incentive: get a real gain, check it in quickly
- From each according to ability
- Data is important; eval data is paramount

Keeping it real

- Working, deployed system that solves a whole problem
- Tight feedback loop
- Everything that matters gets measured

Pedestrian approaches versus cutting edge

- Translate: world-beating and obsolete
- Data versus Syntax
- Language modeling: “Stupid Backoff” (Brants et al., 2007)
- When and how to invest in promising researchy approaches?

Reward and recognition

- Cleverness, independence, origination of new ideas?
- Cooperation, generosity, communication, productivity, risk taking?
- Imposter syndrome
- Happiness

Summary: what's worked for me?

- Work on real systems
- Measure what matters
- Incent the right things
- Keep aware of new research while investing conservatively

Then and now



Thank you!