

The History of Optical Character Recognition and the Google Books Archive*

*A unique title and topic that I came up with

Optical Character Recognition (OCR)

The electronic or mechanical conversion of images of typed, handwritten or printed text into machine-encoded text, whether from a scanned document, a photo of a document, a scene-photo¹

Early optical character recognition may be traced to technologies involving telegraphy and creating reading devices for accessibility.

1. https://en.wikipedia.org/wiki/Optical_character_recognition

1917

The Optophone

E. E. Fournier d'Albe patents the Optophone.

Light from a printed page reflects on a selenium cell, and the machine lets out a musical chord.²



A man demonstrates the Optophone circa. 1921.

2. <https://medium.com/ocrology/a-quick-history-of-optical-character-recognition-ocr-c916d58e2170>

1917

The Optophone.

E. E. Fournier d'Albe patents the Optophone.

Light from a printed page reflects on a selenium cell, and the machine lets out a musical chord.²

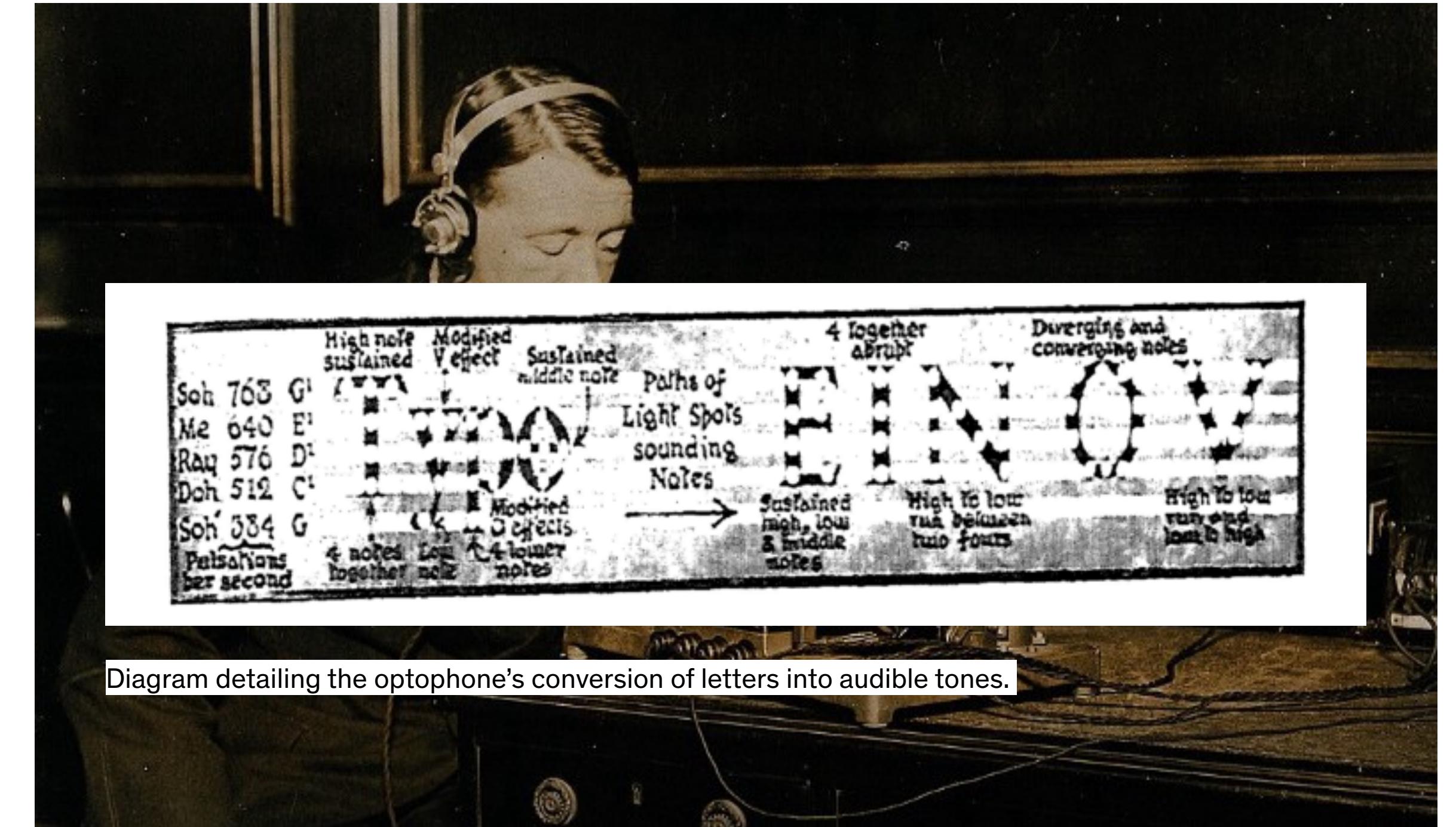


Diagram detailing the optophone's conversion of letters into audible tones.

2. <https://medium.com/ocrology/a-quick-history-of-optical-character-recognition-ocr-c916d58e2170>

1917

The Optophone.

E. E
Light
cell,

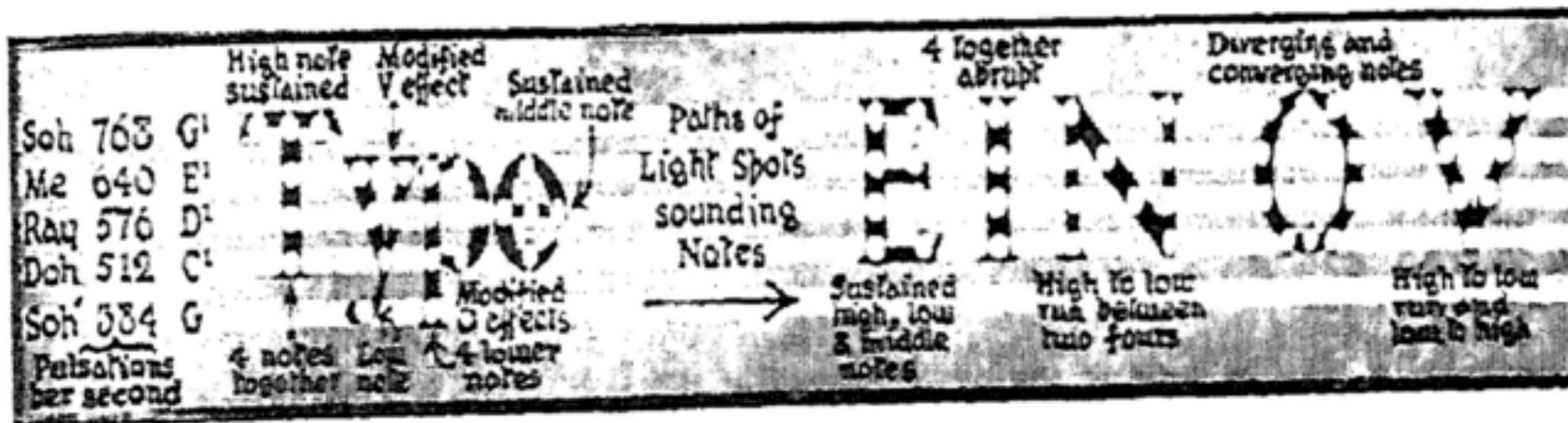


Diagram detailing the optophone's conversion of letters into audible tones.

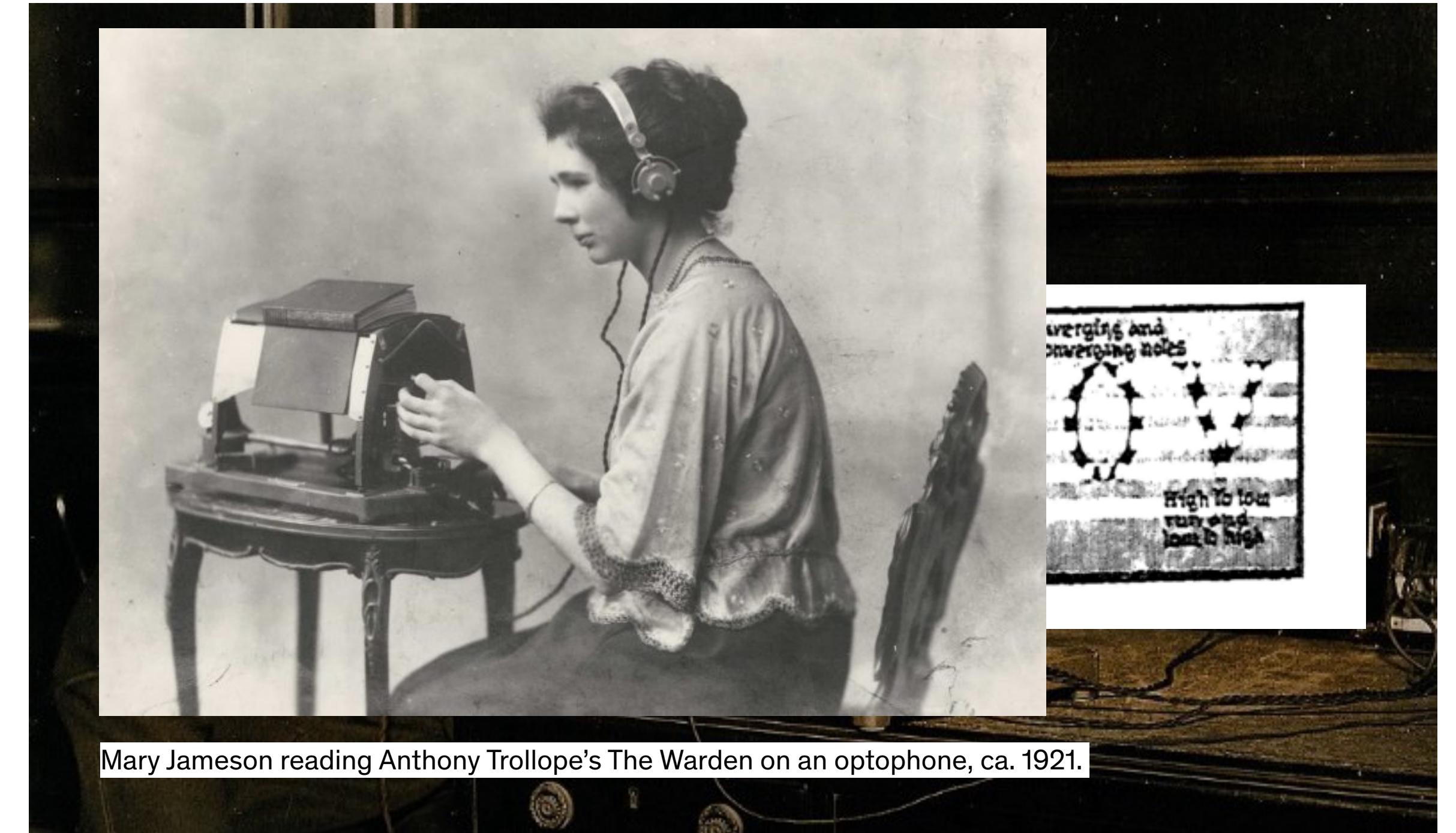
1917

The Optophone.

E. E. Fournier d'Albe patents the Optophone.

Light from a printed page reflects on a selenium cell, and the machine lets out a musical chord.

Mary Jameson read the print material at a record-breaking speed of one word per minute.²

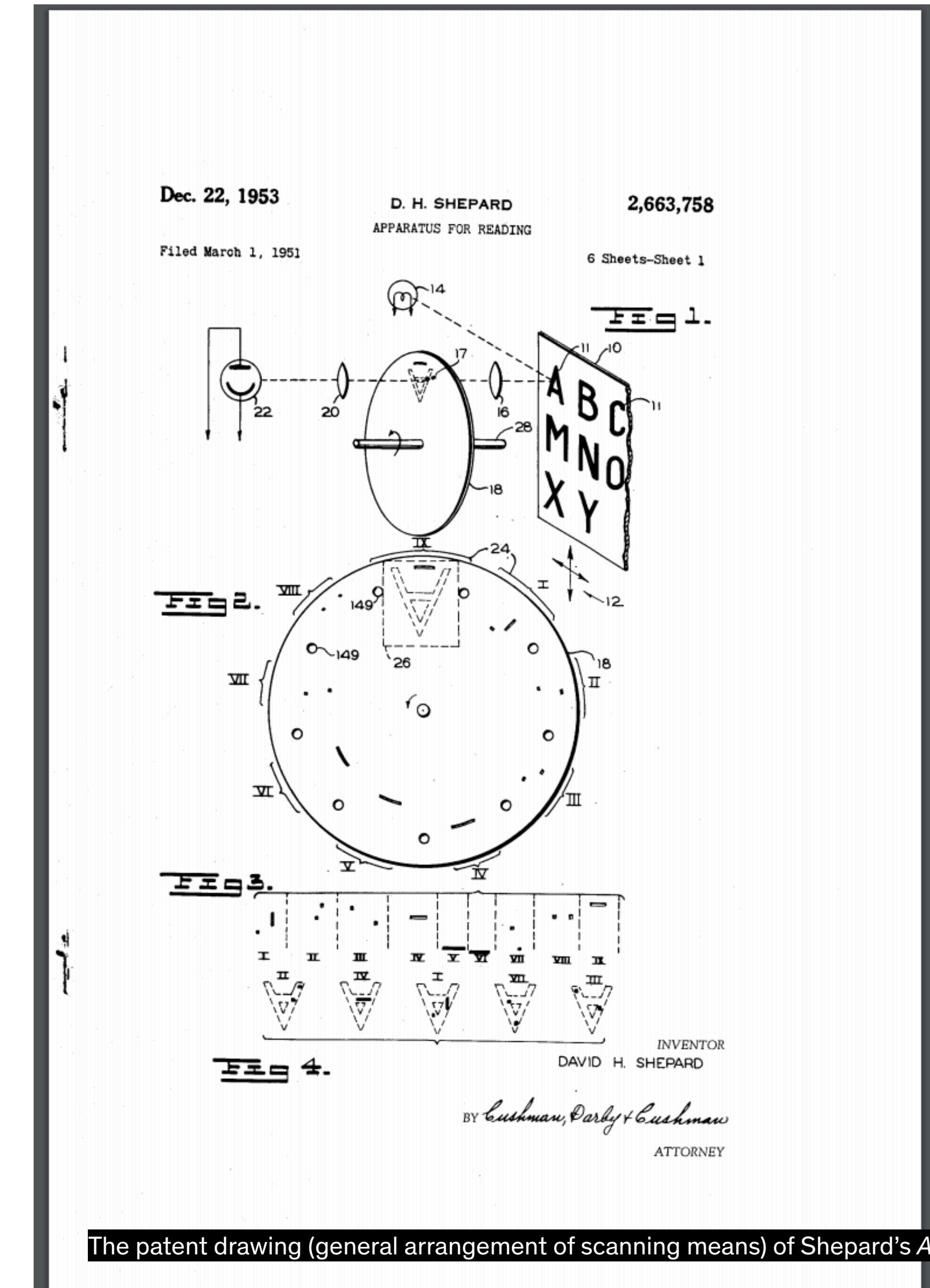


Mary Jameson reading Anthony Trollope's The Warden on an optophone, ca. 1921.

2. <https://medium.com/ocrology/a-quick-history-of-optical-character-recognition-ocr-c916d58e2170>

1951 Gismo

David Shepard develops Gismo, a machine that could recognize all 26 letters of the Latin alphabet, as produced by a standard typewriter. The device later evolved into the Farrington Machine. By the 1960s, OCR technology is being used en-masse in mail-sorting by the U.S. postal service.²



The patent drawing (general arrangement of scanning means) of Shepard's *Apparatus for Reading*

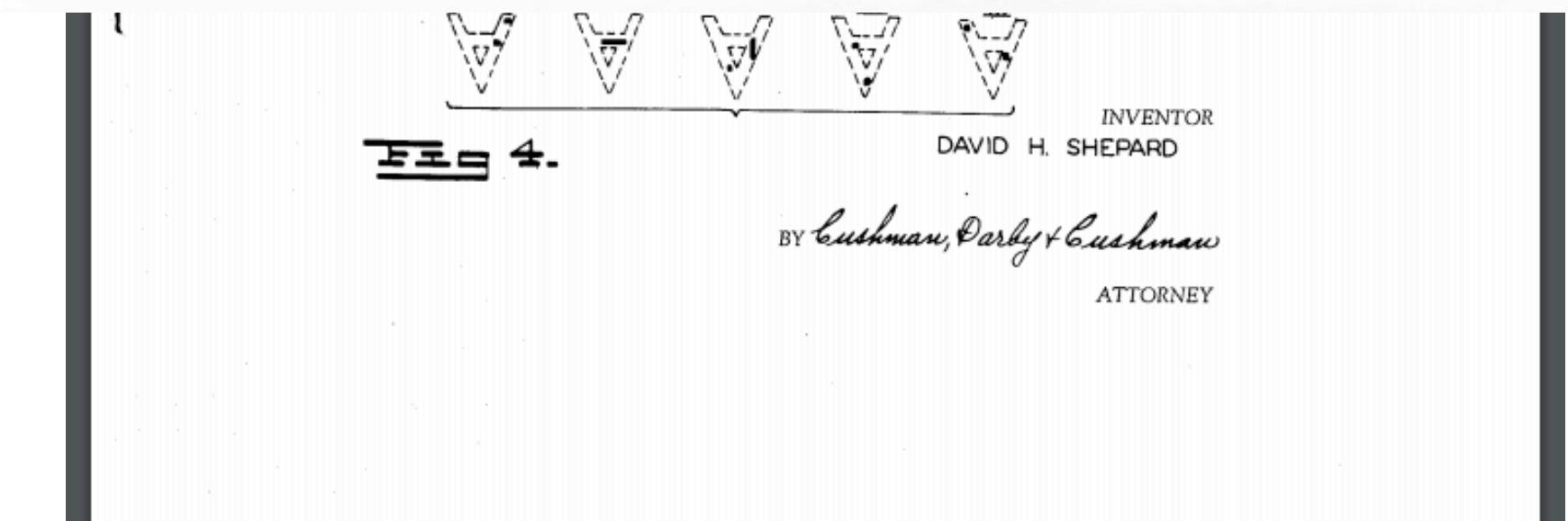
2. <https://medium.com/ocrology/a-quick-history-of-optical-character-recognition-ocr-c916d58e2170>

1951 Gismo

David Shepard develops Gismo, a machine that could recognize all 26 letters of the Latin alphabet, as produced by a standard typewriter. The device later evolved into the Farrington Machine. By the 1960s, OCR technology is being used en-masse in mail-sorting by the U.S. postal service.²



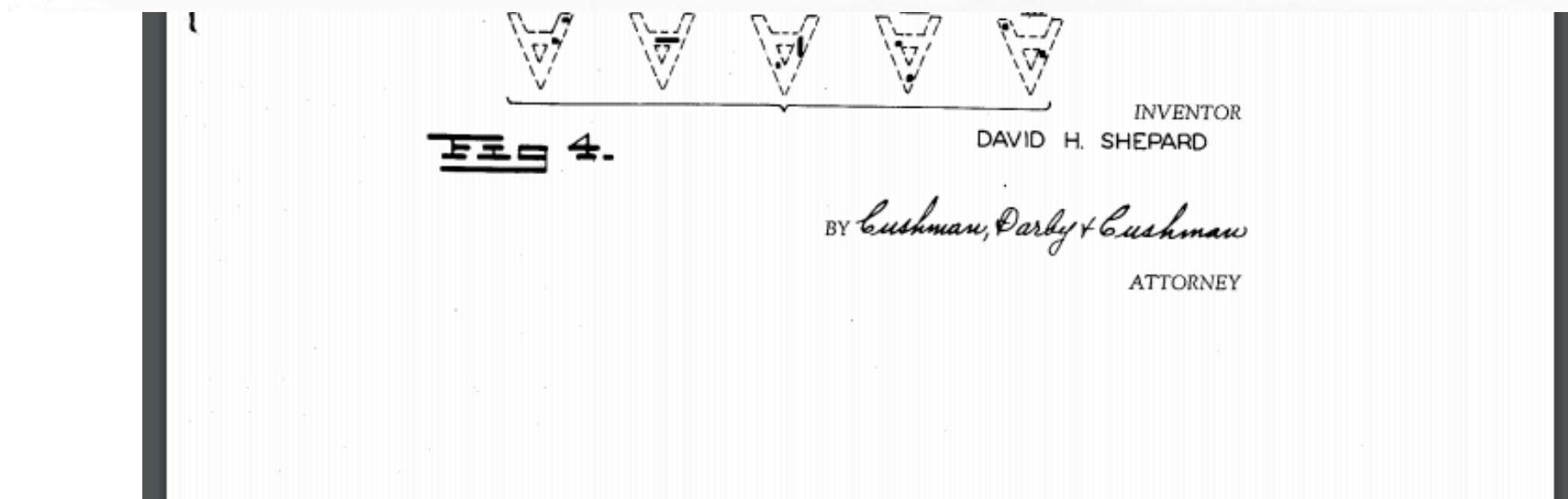
Postal employee showing off the Farrington Automatic Address Reader in the 1960s.



2. <https://medium.com/ocrology/a-quick-history-of-optical-character-recognition-ocr-c916d58e2170>

1951 Gismo

David Shepard develops Gismo, a machine that could recognize all 26 letters of the Latin alphabet, as produced by a standard typewriter. The device later evolved into the Farrington Machine. By the 1960s, OCR technology is being used en-masse in mail-sorting by the U.S. postal service.²



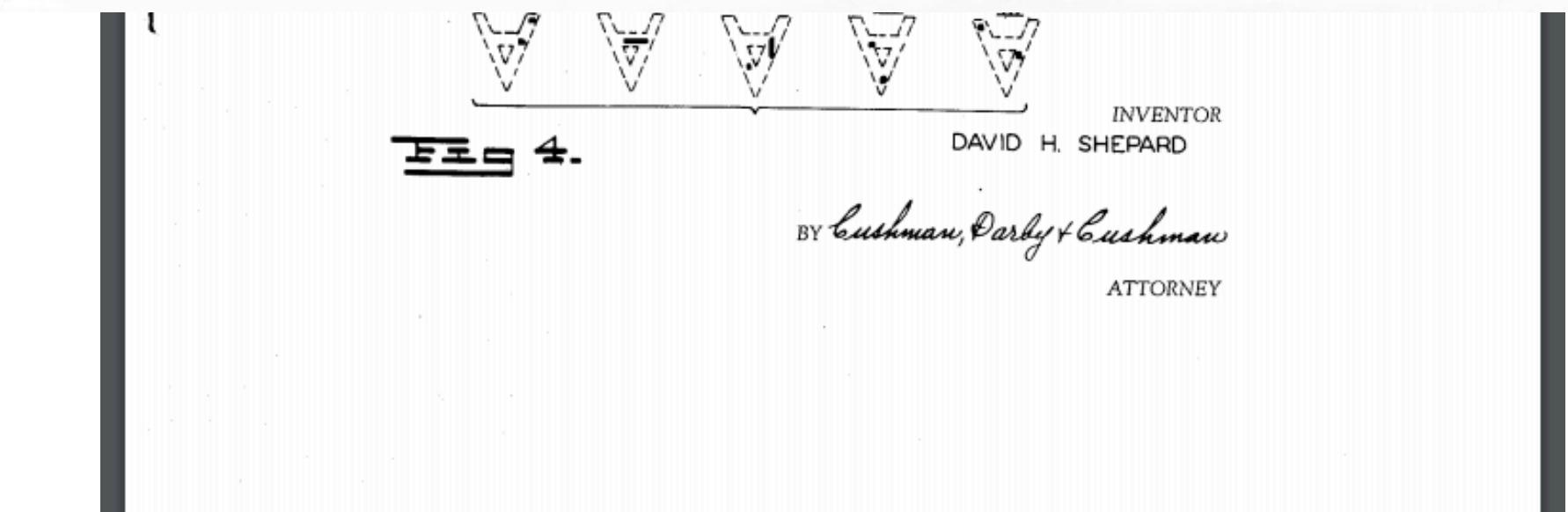
2. <https://medium.com/ocrology/a-quick-history-of-optical-character-recognition-ocr-c916d58e2170>

1951 Gismo

David Shepard develops Gismo, a machine that could recognize all 26 letters of the Latin alphabet, as produced by a standard typewriter. The device later evolved into the Farrington Machine. By the 1960s, OCR technology is being used en-masse in mail-sorting by the U.S. postal service.²



Farrington 7B font is developed to support clear transmission using analog card processing



2. <https://medium.com/ocrology/a-quick-history-of-optical-character-recognition-ocr-c916d58e2170>

1974 CCD flatbed scanner

Kurzweil Technologies releases the CCD flatbed scanner, the first omni-font optical character recognition system. Before 1974, OCR could only read fonts specially designed for machine readability.²



2. <https://medium.com/ocrology/a-quick-history-of-optical-character-recognition-ocr-c916d58e2170>

1992

Newton MessagePad

Kurzweil Technologies releases the CCD flatbed scanner, the first omni-font optical character recognition system. Before 1974, OCR could only read fonts specially designed for machine readability.



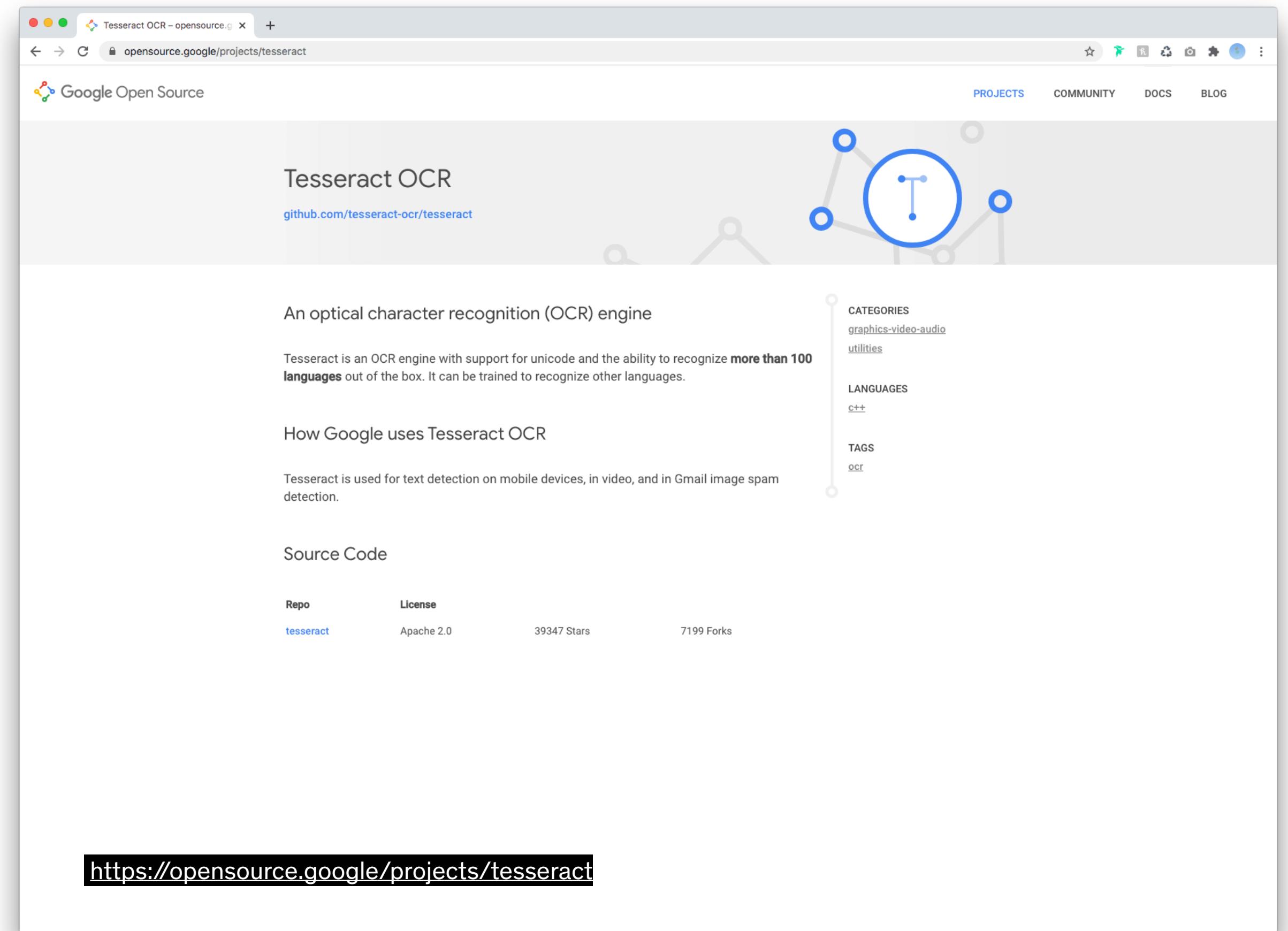
A Newton MessagePad 2100 (circa 1997, left) and iPhone 11 Pro Max (right).

2. <https://medium.com/ocrology/a-quick-history-of-optical-character-recognition-ocr-c916d58e2170>

1985–Present

Tesseract

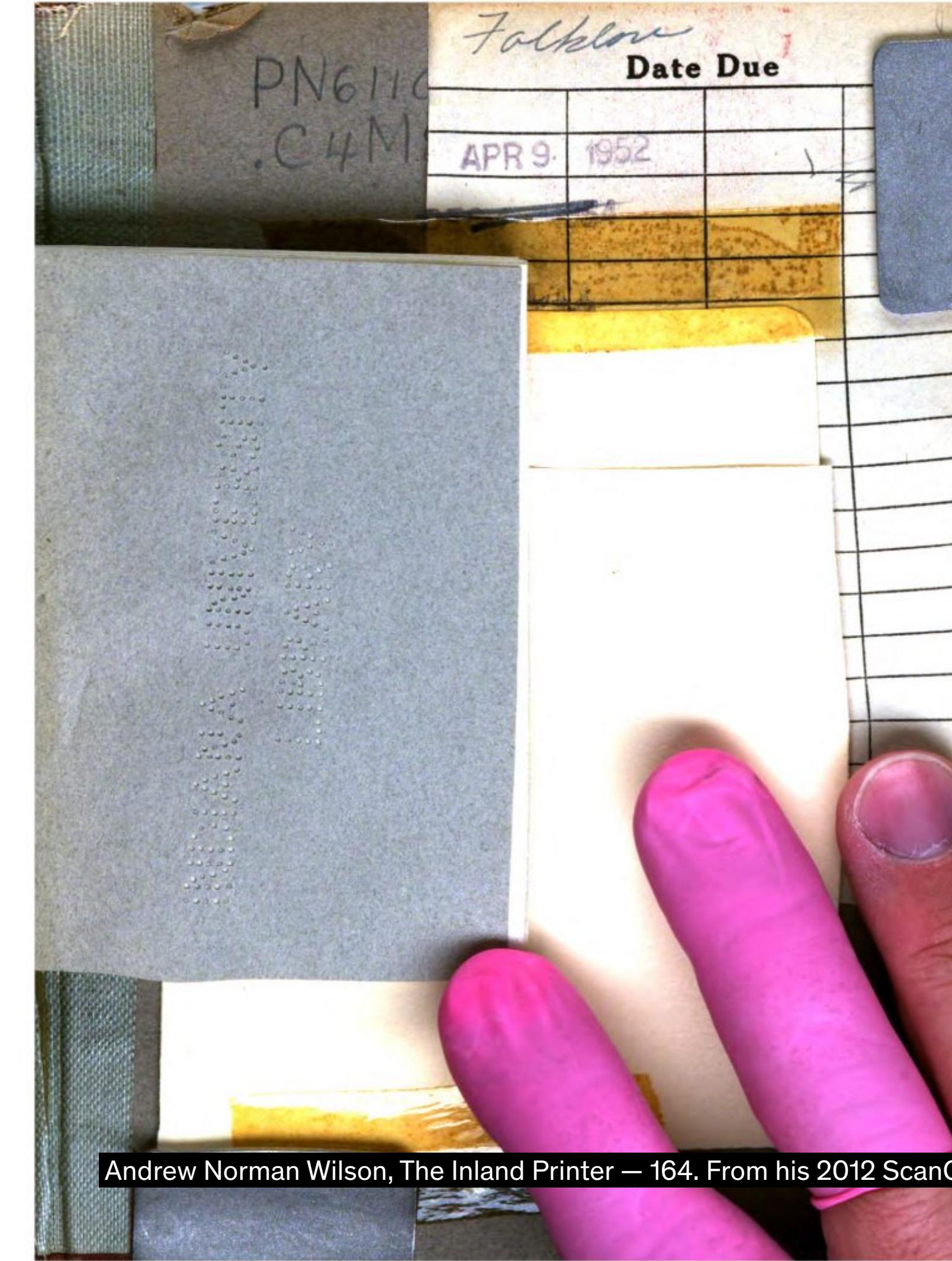
In 2006, Google takes the OCR software, Tesseract, under its wing, accelerating industry collaboration. In the coming years, OCR is married to neural networks. Instead of having to input the rules of language into a machine, neural networks will allow OCR to recognize patterns for themselves.



2005 Google's Universal Library

Empowered by Tesseract, Google embarks on a project to digitally archive “every book on the planet in every language.”

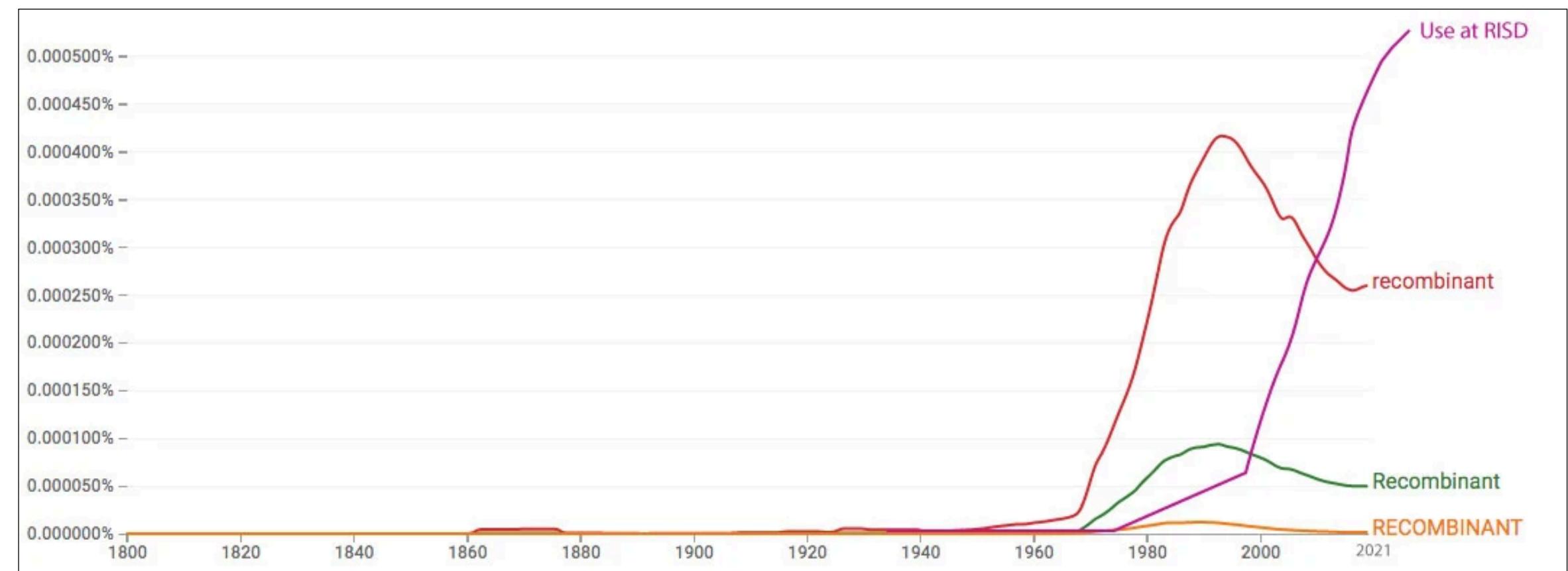
- Project Ocean
- “Library of Utopia”
- Public vs. Private



Books Ngram Viewer

The Google Ngram Viewer or Google Books Ngram Viewer is an online search engine that charts the frequencies of any set of search strings using a yearly count of n-grams found in sources printed between 1500 and 2019.

An n-gram is a contiguous sequence of n items from a given sample of text or speech



3. https://en.wikipedia.org/wiki/Google_Books

Google Scanning

As of October 2015, the number of scanned book titles was over 25 million in more than 400 languages.

Infrared camera technology that detects the three-dimensional shape and angle of book pages

The vast majority of books scanned are in English, resulting in a disproportionate representation of natural languages and knowledge in the digital world.

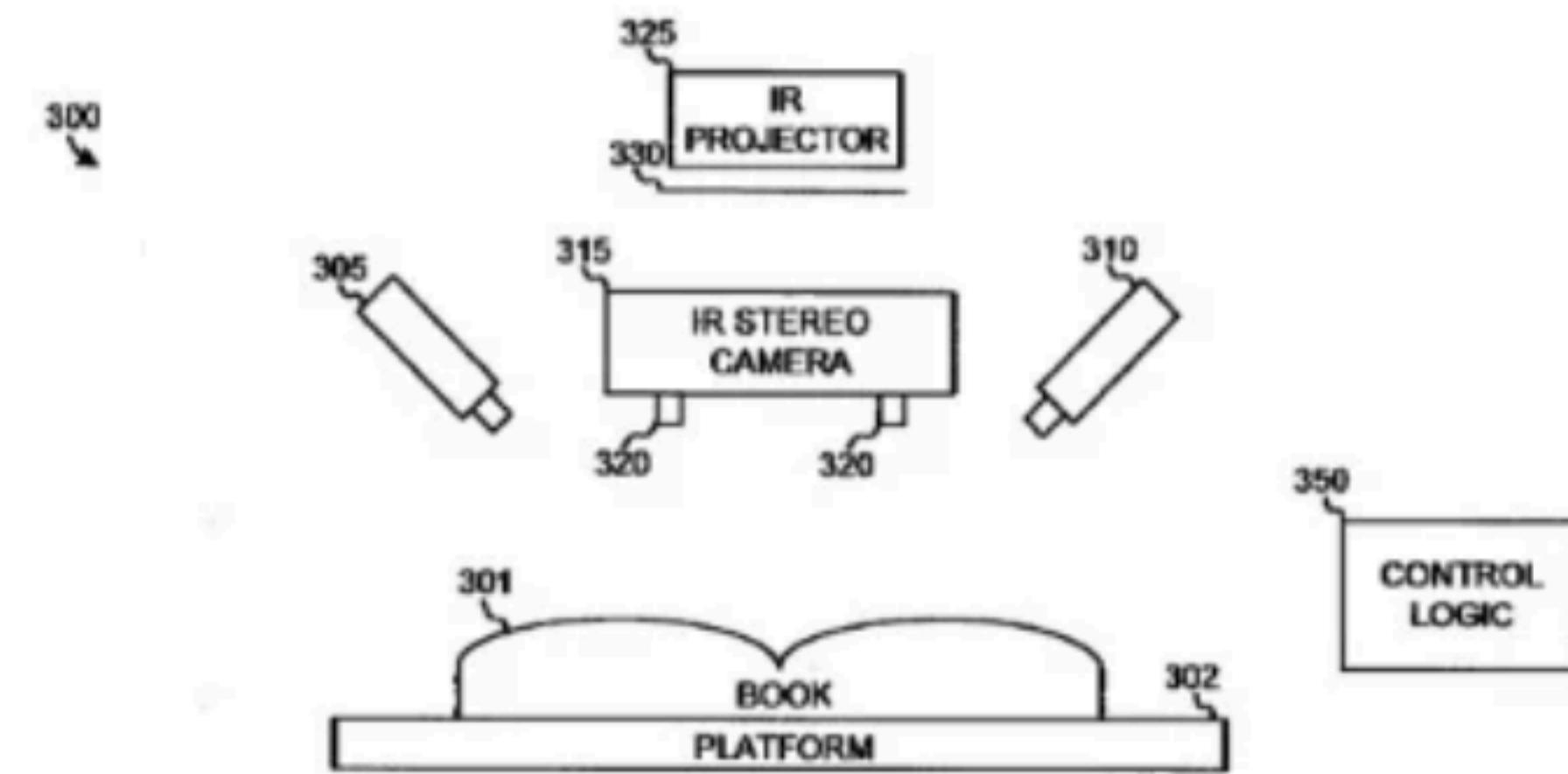


FIG. 3

Image of Google's infrared camera technology, [Patent 7508978](#)

Preserving Local Writers, Genealogy, Photographs, Newspapers, and Related Materials

books.google.com/books?id=_MIVfCA2QUC&pg=PA311&lpg=PA311&dq="The+archive+is+a+site+of+ambiguity.+It+is+best+understood+as+a+contested+terrain+for+memory+construction+that+in+turn+shapes+contemporary+understanding+of+society."&pg=PA311&printsec=frontcover

Google Books

VIEW EBOOK

Get this book in print ▾

★★★★★ 0 Reviews Write review

Preserving Local Writers, Genealogy, Photographs, Newspapers, and Related Materials

edited by Carol Smallwood, Elaine Williams

Search in this book Go

About this book

My library

My History

Books on Google Play

formers integrated into vaudeville, but stereotypes and prejudice remained. By the end of vaudeville and into the world wars, coon songs and the minstrelsy were still alive and were considered lowbrow comedy. It caused problems during the Second World War. For example, the Hollywood cartoon *Coal Black and de Sebben Dwarfs*, with its racist caricaturing of African American troops, was banned, owing to the protests of the NAACP and concerns of the U.S. war effort that it would lower the morale of troops.⁸

Purpose

The archive is a site of ambiguity. It is best understood as a contested terrain for memory construction that in turn shapes contemporary understanding of society.

—Report, Nelson Mandela Foundation, 2005⁹

Collections that include offensive sheet music should be prefaced with carefully worded introductions, depending on the decisions of local archivists, in adherence to professional ethical codes of impartiality. They should provide context, but should not draw too much attention to individual items. When one comes upon a racist item (a sheet music cover page with a racist caricature, a title with a slur, racist lyrics, etc.), one's inclination is to separate it; one subconsciously places emphasis—special significance—to the item. However, the collection as a

In the Dream House - Google Books

Table of contents unava... "What is placed in or left out of the archive is a political act, dictated by the archivist and the political context in which she lives."

Result 1 of 1 in this book for "What is placed in or left out of the archive is a political act, dictated by the archivist and the political context in which she lives."

By Carmen Maria Machado

Preview

Overview Get 1-day借阅

About this edition

ISBN: 978164413000
Published: November 2018
Publisher: Graywolf Press
Author: Carmen Maria Machado

Create Citation

A revolutionary memoir about family, love, and parties.

In the Dream House is Carmen Maria Machado's powerful memoir of her relationship gone bad, and her struggle against psychological abuse. Traumatized by her mother's abuse, Machado struggled to come to terms with her own becoming.

And it's that struggle that informs Machado's writing, a trope—the haunted house—that she uses to explore the light and examines the dark.

Source: Publisher

Dream House as Prologue

In her essay “Venus in Two Acts,” on the dearth of contemporaneous African accounts of slavery, Saidiya Hartman talks about the “violence of the archive.” This concept—also called “archival silence”—illustrates a difficult truth: sometimes stories are destroyed, and sometimes they are never uttered in the first place; either way something very large is irrevocably missing from our collective histories.

The word *archive*, Jacques Derrida tells us, comes from the ancient Greek ἀρχεῖον: *arkheion*, “the house of the ruler.” When I first learned about this etymology, I was taken with the use of *house* (a lover of haunted house stories, I’m a sucker for architecture metaphors), but it is the power, the authority, that is the most telling element. What is placed in or left out of the archive is a political act, dictated by the archivist and the political context in which she lives. This is true whether it’s a parent deciding what’s worth recording of a child’s early life or—like Europe and its *Stolpersteine*, its “stumbling blocks”—a continent publicly reckoning with its past. *Here is where Sebastian took his first fat-footed baby steps; here is the house where Judith was living when we took her to her death.*

Sometimes the proof is never committed to the archive—it is not considered important enough to record, or if it is, not important enough to preserve. Sometimes there is a

Clear search

5, 2019
autobiography
graphy / LGBT, more

Carmen Maria Machado
The author of *Her Body and Other Parties* and finalist for the National Book Award, Machado will speak at the National Book Critics Circle's annual awards dinner in Philadelphia with her fellow winners.

Machado

Her Body and Other Parties

Scan Ops

Andrew Norman Wilson



Scan Ops

Andrew Norman Wilson



Scan Ops

Andrew Norman Wilson



Scan Ops

Andrew Norman Wilson



Cruising Utopia, 10th Anniversary

books.google.com/books?id=Zy2RDwAAQBAJ&pg=PA65&lpg=PA65&dq=*Jacques+Derrida's+idea+of+the+trace+is+relevant+here.+Ephemeral+evidence+is+rarely+obvious+because+it+is+needed+to+stand+against+... Page 65

Google *Jacques Derrida's idea of the trace is relevant here. Ephemeral evidence Try the new Google Books >

Books

VIEW EBOOK

Result 1 of 1 in this book for *Jacques Derrida's idea of the trace is relevant here. Ephemeral evidence is rarely obvious because it is needed to stand against the harsh lights of mainstream visibility and the potential tyranny of the fact.

Clear search

Get this book in print ▾

★★★★☆ 0 Reviews Write review

Cruising Utopia, 10th Anniversary
Edition: The Then and There of Queer Futurity
By José Esteban Muñoz, Joshua Chambers-Letson, Tavia Nyong'o, Ann Pellegrini

*Jacques Derrida's idea Go

About this book
My library
My History
Books on Google Play

NYU PRESS

Pages displayed by permission of NYU Press. Copyright.

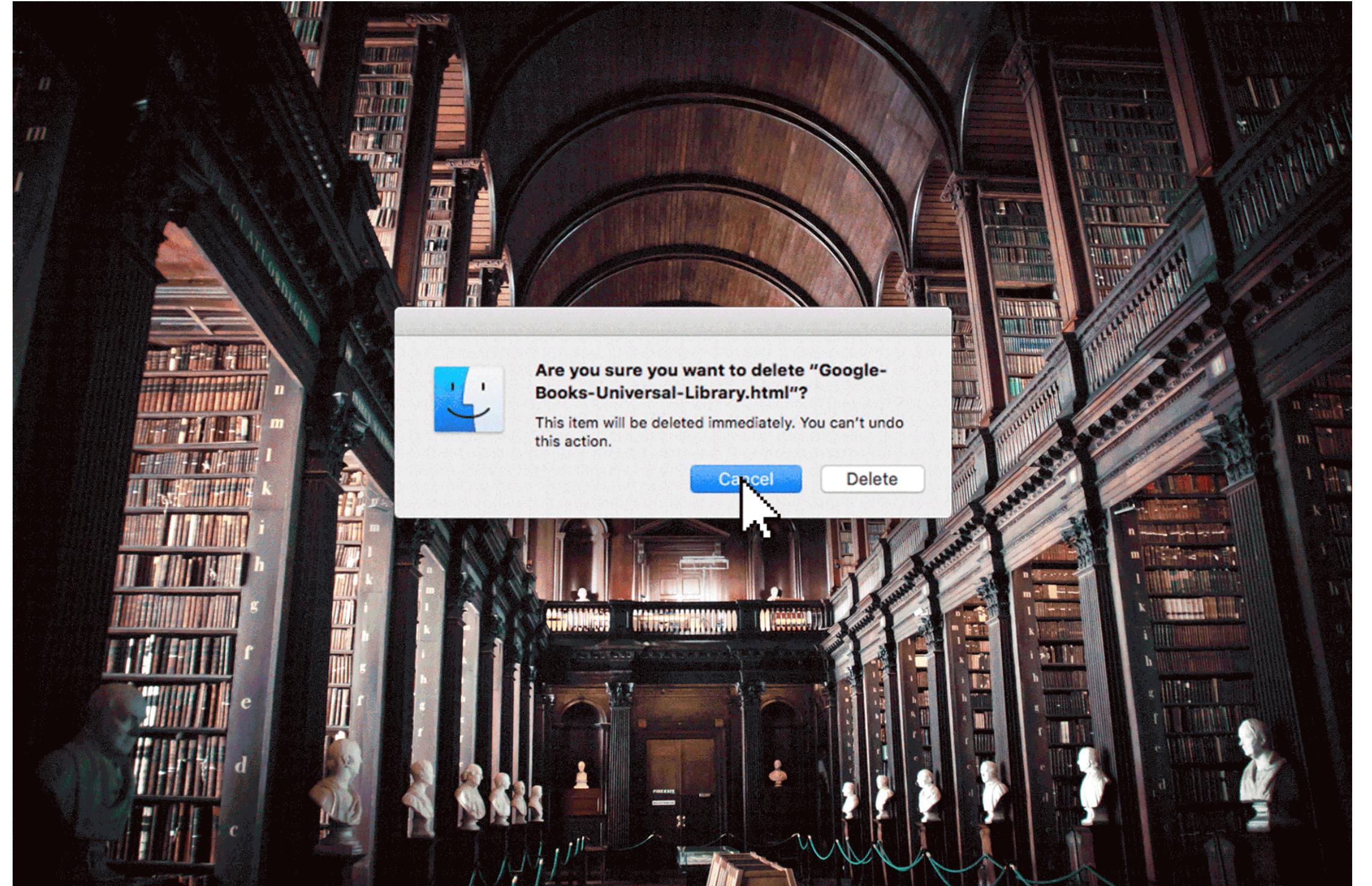
THIS CHAPTER HAS two beginnings.¹ One is a story culled from personal memory, and the other is a poem by a prominent twentieth-century North American poet. Both openings function as queer evidence: an evidence that has been queered in relation to the laws of what counts as proof. Queerness has an especially vexed relationship to evidence. Historically, evidence of queerness has been used to penalize and discipline queer desires, connections, and acts. When the historian of queer experience attempts to document a queer past, there is often a gatekeeper, representing a straight present, who will labor to invalidate the historical fact of queer lives—present, past, and future. Queerness is rarely complemented by evidence, or at least by traditional understandings of the term. The key to queering evidence, and by that I mean the ways in which we prove queerness and read queerness, is by suturing it to the concept of ephemera. Think of ephemera as trace, the remains, the things that are left, hanging in the air like a rumor.

Jacques Derrida's idea of the trace is relevant here.² Ephemeral evidence is rarely obvious because it is needed to stand against the harsh lights of mainstream visibility and the potential tyranny of the fact. (Not that all facts are harmful, but the discourse of the fact has often cast antinormative desire as the bad object.) Ephemera are the remains that are often embedded in queer acts, in both stories we tell one another and communicative physical gestures such as the cool look of a street cruise, a lingering handshake between recent acquaintances, or the mannish strut of a particularly confident woman.

In this chapter I want to approach the idea of queerness and gesture. So much can be located in the gesture. Gesture, I argue throughout this book, signals a refusal of a certain kind of finitude. Dance is an especially valuable site for ruminations on queerness and gesture. This theoretical work is anchored to a case study, a living body, a performer who is a master of

Legal Challenges

Through the project, library books were being digitized somewhat indiscriminately regardless of copyright status, which led to a number of lawsuits against Google. By the end of 2008, Google had reportedly digitized over seven million books, of which only about one million were works in the public domain. The Authors Guild and Association of American Publishers separately sued Google in 2005 for its book project, citing "massive copyright infringement."



<https://www.theatlantic.com/technology/archive/2017/04/the-tragedy-of-google-books/523320/>

2. <https://medium.com/ocrology/a-quick-history-of-optical-character-recognition-ocr-c916d58e2170>

ZXX

Sang Mun

A B C

~~BELOWING~~
~~THE~~
~~WITNESSHTLE~~
~~ON A WAR~~
~~XCRIME~~
~~SHIELD~~
~~NOT ON A~~
~~CREME~~

Redaction

redaction

