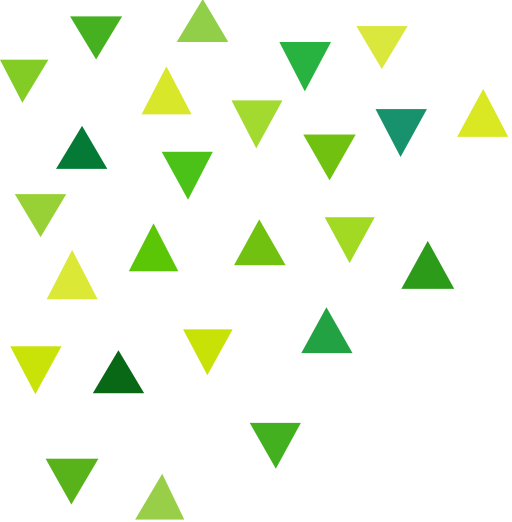


Cannlytics

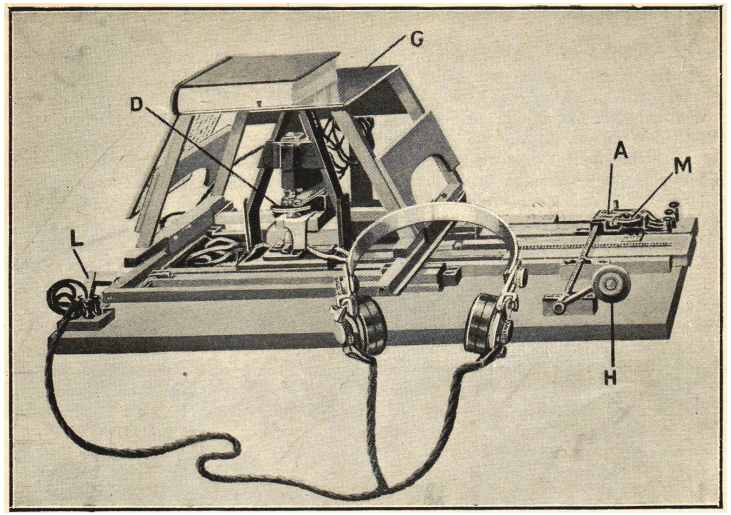
Cannabis Data Science #82

September 7th, 2022

Early Optical Character Recognition (OCR)

- 1914 - 1931: **Emanuel Goldberg** created machines to read characters and convert them into standard telegraph code.
- 1974: **Ray Kurzweil** monetized OCR to recognize virtually any printed font.

Photo by Michael Lutch
License: CC BY 1.0
https://creativecommons.org/licenses/by/1.0



An **optophone**, scans text and generates time-varying chords of tones to identify letters.

Note: Uses selenium photosensors!

- 1978: **LexisNexis** parsed legal documents into online databases.
- 2000s: First web-based OCR applications.

Cannabis Data Science Application

Questions of the Day

- Can the data be extracted from a **COA** image taken by a consumer at a retailer?
- Given consumer's history, $\overline{X}$, can we predict product(s) with chemical profile(s), $\hat{Y}$, that the consumer may *enjoy*?

Product Recommendation Algorithm

Algorithm:

1. Calculate average product concentrations by consumer, $\bar{X}$;
2. Find k-nearest products, $\hat{Y}$, given the average concentrations;
3. Repeat when the consumer makes a purchase.

Extension:

- Weighted average by **sentiment analysis**:

$$\bar{X} = \frac{\sum_{m=1}^{M} \theta_m X_m}{\sum_{m=1}^{M} \theta_m},$$

where θ_m is a consumer's sentiment score (0 to 1) for a review of each product, m, of M total products.

 Thank you for coming.

Insights of the Day

- Start somewhere, then **iterate**.
- The place where you begin **doesn't have to be perfect**.

What is on your mind for next week?