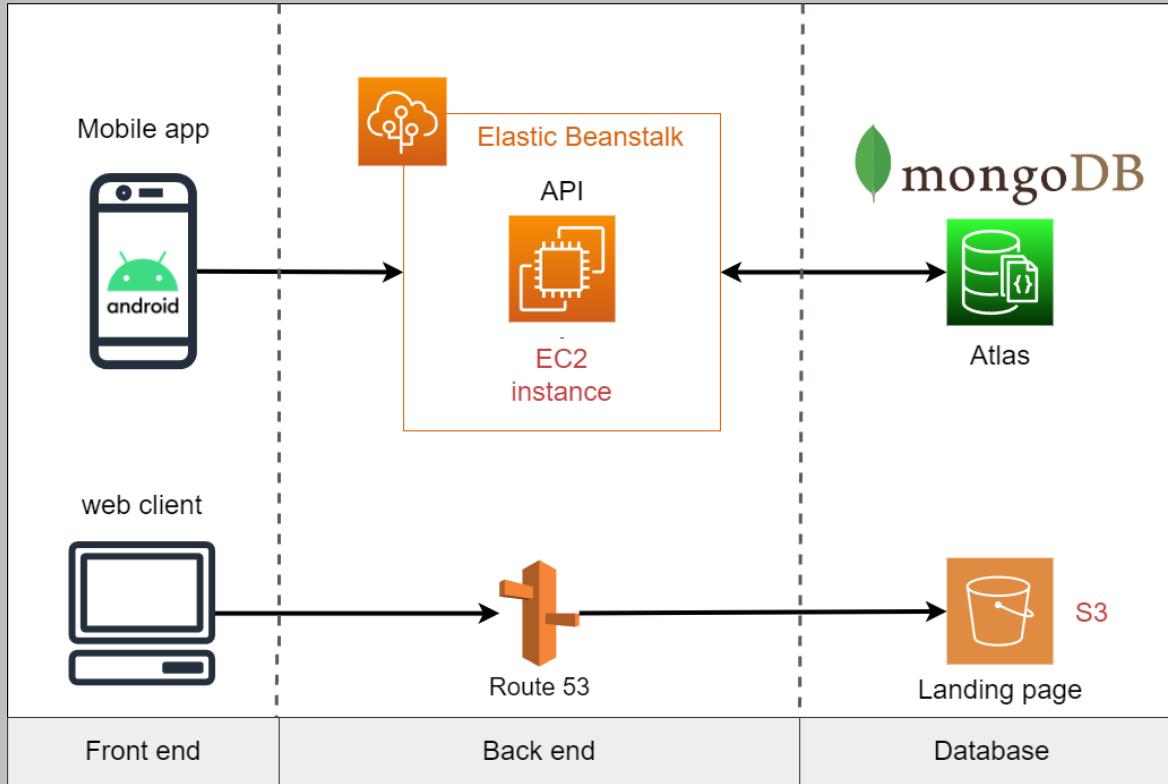




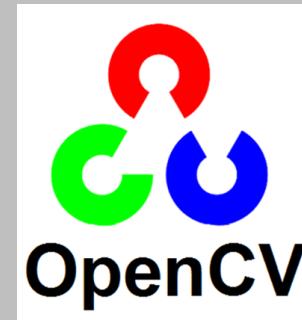
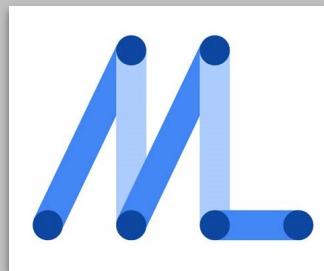
Scanalyze

SPEND SMART

Architecture



Application android



MPAndroidChart

created by Philipp Jahoda

Scan workflow

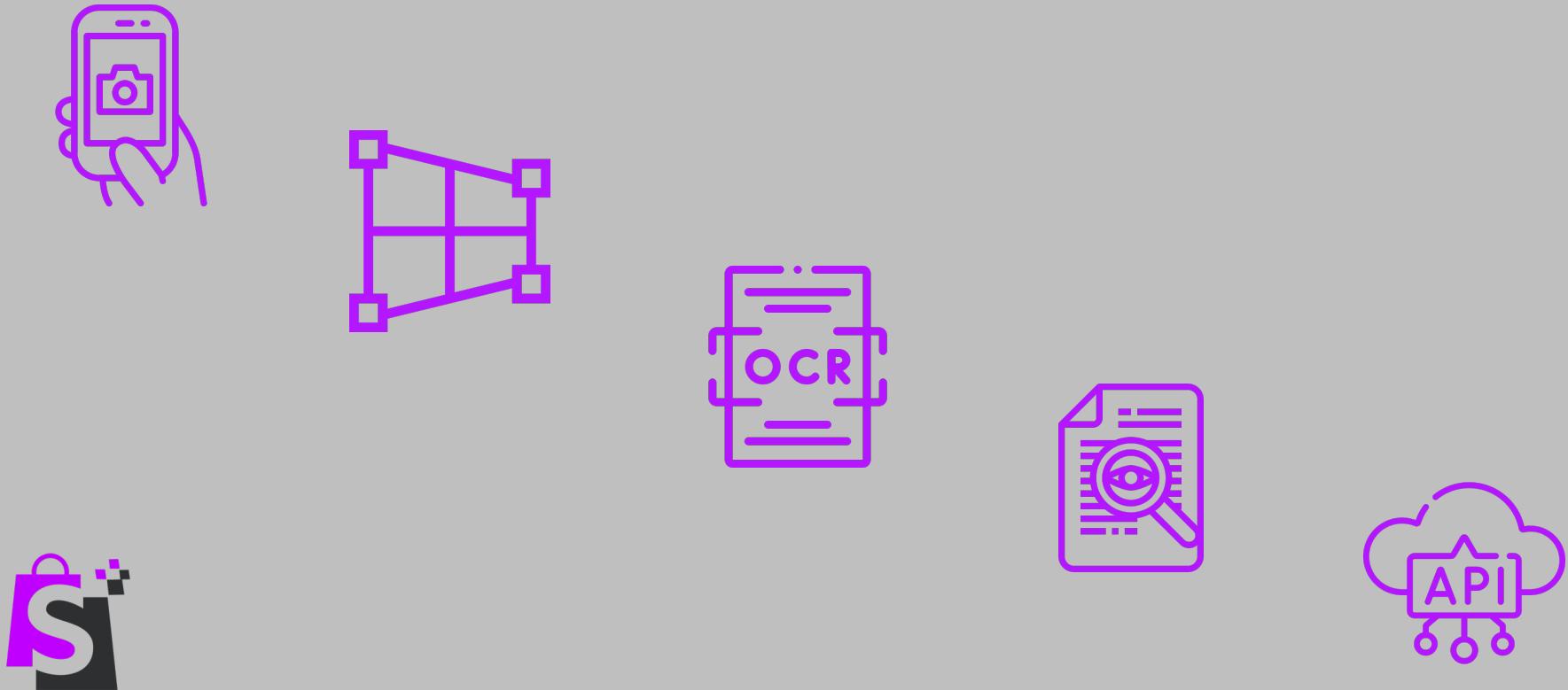


Image pre-processing

Image source



Grey scale

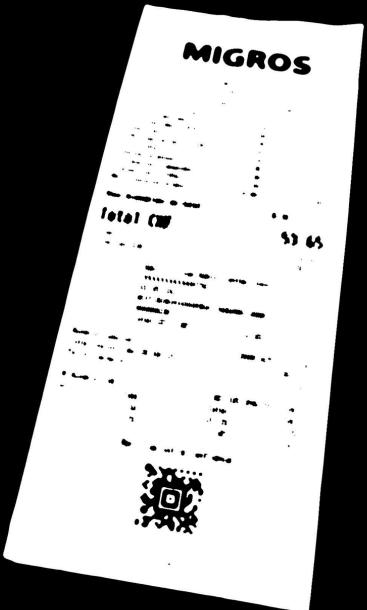


Gaussian blur

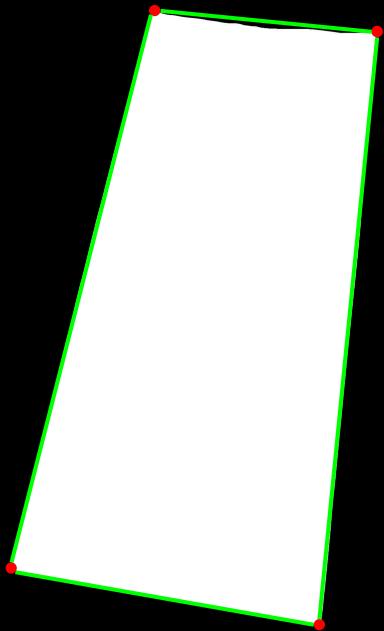


Image pre-processing

OTSU Thresholding



Dilation / Erosion



Application à image source



Correction de perspective et post-processing



Détection OCR

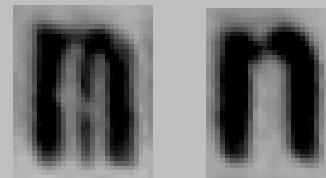
- Blocs -> groupe de texte
- Lignes -> lignes de mots
- Elements -> mots



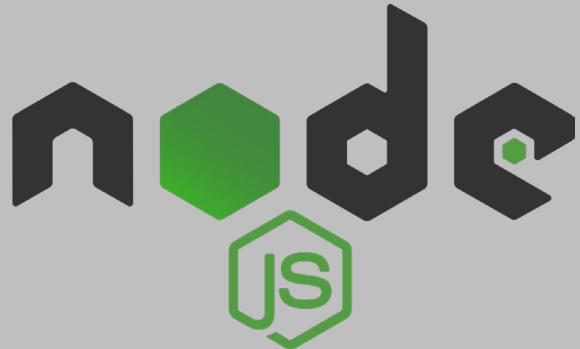
Reconstruction des lignes

- Éléments triés par coordonnée Y des bounding boxes
- reconstruction des lignes par coordonnée X
- Utilisation de regex pour détecter les patterns

MCClass	pois	chiches	1	1.20	1.20	1
MCClass	bigarreaux	235g	2	2.50	5.00	1
Elmex	eau	dent. Duo	2	13.00	8.10	17.90
Pommes Jazz			1	4.95	5.25	2
Tonka						1



API - Technologies



express



API - Technologies

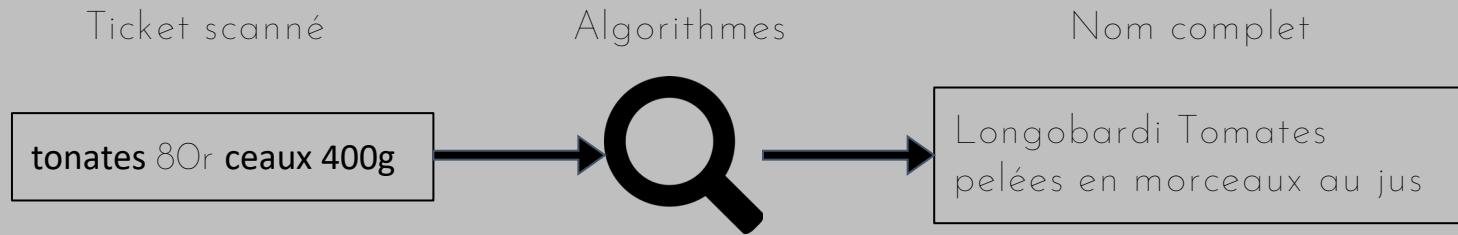


API - Algorithmes d'indexation

Coefficient de dice	Distance de Levenshtein	Distance de Jaro-Winkler
<ul style="list-style-type: none">• Division en n-gram• Comparaison de n-gram similaire	<ul style="list-style-type: none">• Coût pour transformer s1 en s2• Substitution / ajout / suppression	<ul style="list-style-type: none">• Caractère correspondant• Nombre de transpositions



API - Indexation



API - Routes

receipts Everything about the receipts ^

GET /receipts Return a list of receipts. ▾

POST /receipts Post a new receipt. ▾

GET /receipts/{id} Return a receipt by id ▾

statistics Access to statistics ^

GET /statistics Return statistics. ▾



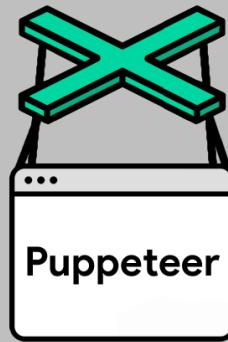
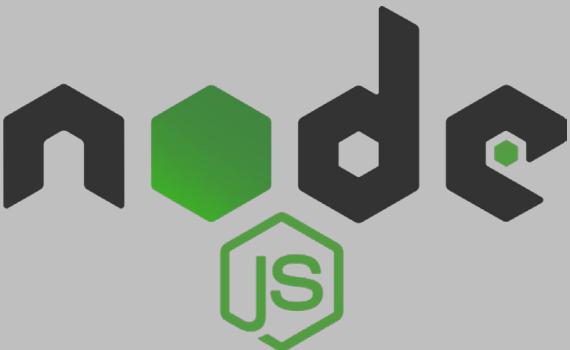
Scrapers - Utilité

- Peuplement de la base
- Pas d'accès légitime
- Données en libre service

The screenshot shows the homepage of the Coop website. At the top, there is a navigation bar with links for "Ma Coop", "Nourriture", "Vins", "Ménage & Animaux", "Cosmétiques & Santé", "Bébé & Enfant", "Marques & Inspirations", "Actions", and "Panier". Below the navigation bar, there is a search bar and links for "Magasins", "Disponible en ligne", and "Connexion". The main content area features a section titled "Points forts de la semaine" (Week's highlights) with four items: "Pruneaux 1.5kg" (4.95 CHF), "Viande de bœuf haché 2x400g" (9.95 CHF), "Carte journalière spéciale" (49.00 CHF), and "Potimarrons" (2.45 CHF). Each item has a small image, a price, a discount percentage, and a "Frais de livraison en su" (Delivery fee included) message. Below this section, there is a link "Tout visualiser" (View all) and a page navigation indicator "1 / 3".



Scrapers - Technologies



Database



 MongoDB.[®]
Atlas



Méthodologie de travail

- Application Mobile - Timothée
- API + Scraper - Jarod
- Database + Scraper - Anthony
- Vidéo - Anthony
- Feature branch respectée
- Mise au point journalière
- Feedback entre membre de l'équipe



Bilan

- Infrastructure fonctionnelle
- Application utilisable
- Énorme apprentissage





Scanalyze

SPEND SMART

www.scanalyze.ch