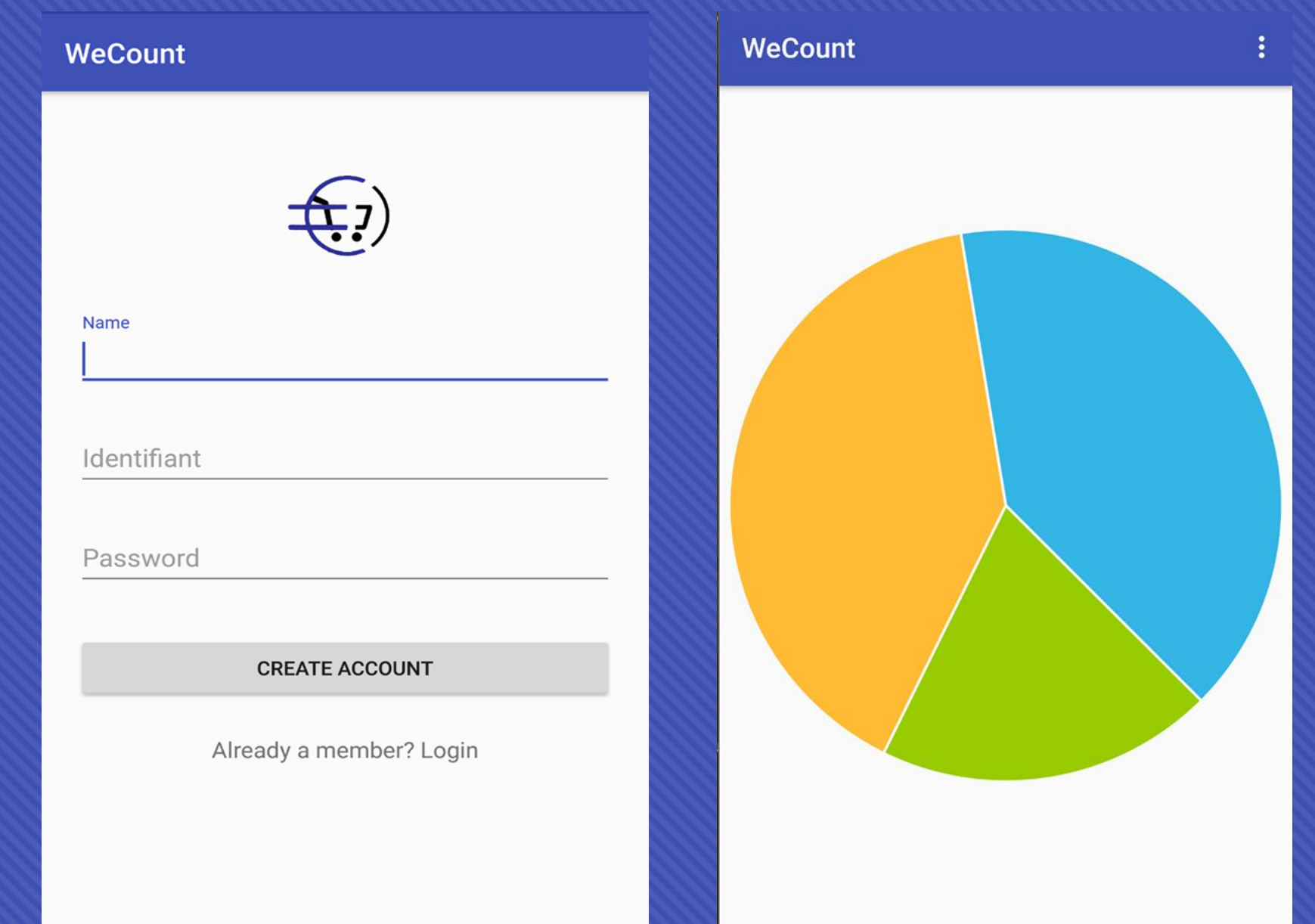




Qui n'a jamais eu de problèmes et n'a jamais passé beaucoup de temps à faire ses comptes avec ses tickets de caisse?

Avec l'application **WeCount** ces problèmes sont derrière vous ! Une simple photo de votre ticket de caisse et votre budget, vos dépenses deviennent plus clairs.

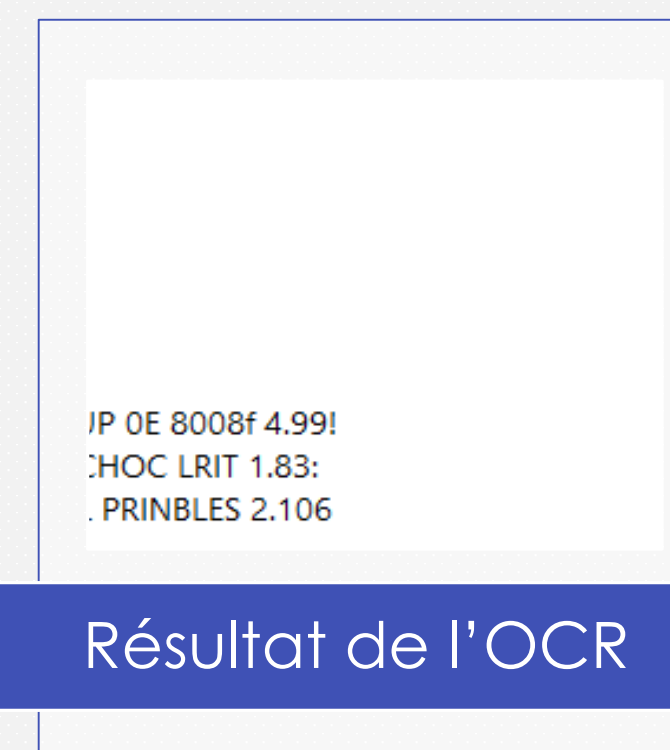
Notre motivation : vous aider à mieux gérer et comprendre vos dépenses.



Ticket de caisse



Méthode d'Otsu



Résultat de l'OCR

IMAGE

Extraire les articles du ticket de caisse et les présenter dans un format texte. On procède par méthode d'Otsu pour binariser l'image, méthode RANSAC pour repérer les bords du ticket, méthode du pivot de Gauss pour appliquer une homographie et on utilise un logiciel OCR au final.

Options	id	Nom_original	id_nom_commun	Prix	id_achat	Quantité
✎	4	1/2PATE CND CRT CH	NULL	2.84	1	1
✎	5	125G MIN FLUTE EMM	NULL	1.5	1	1
✎	6	12X OEUFS MOYENS	5	2.04	1	1
✎	7	150G COMTE RAPE	NULL	1.63	1	1
✎	8	150G TRCHE ROQUEF	NULL	2.58	1	1
✎	9	180G CAROTTE RAPEE	5	2	1	1
✎	10	1KG COUILLETTE P	NULL	2.04	1	1
✎	11	1L HLE OLIVE VE CR	NULL	5.7	1	1
✎	12	20 OEUFS MOYEN DATE	NULL	3.23	1	1
✎	17	Baguette	9	0.9	9	2
✎	18	Baguette	9	0.9	10	2
✎	19	JPOM	1	2.52	13	2

BASE DE DONNEES

- Création d'une BDD SQL en fonction des besoins de l'application et codage en Java sous Eclipse des fonctions permettant aux autres modules de communiquer avec elle.

WEB DES DONNEES

- Remplissage de la BDD grâce au crawl de auchandirect.fr. Récupération des produits grâce à la connaissance de la structure du code html.
- Récupération des promotions sur internet grâce à un crawler (suivi des lien par champ lexical, détection d'un produit au symbole "€", évaluation des pages par intersection de leur contenu avec la BDD).

SES

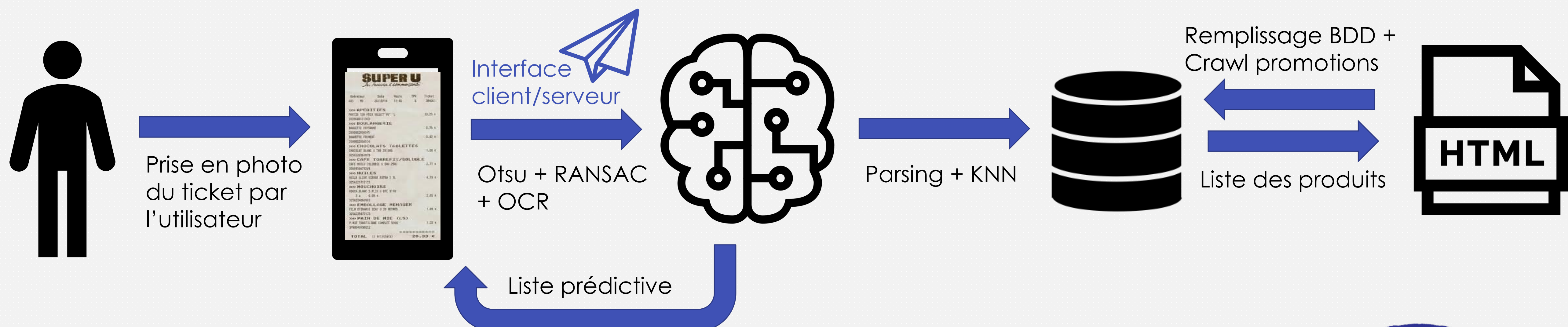
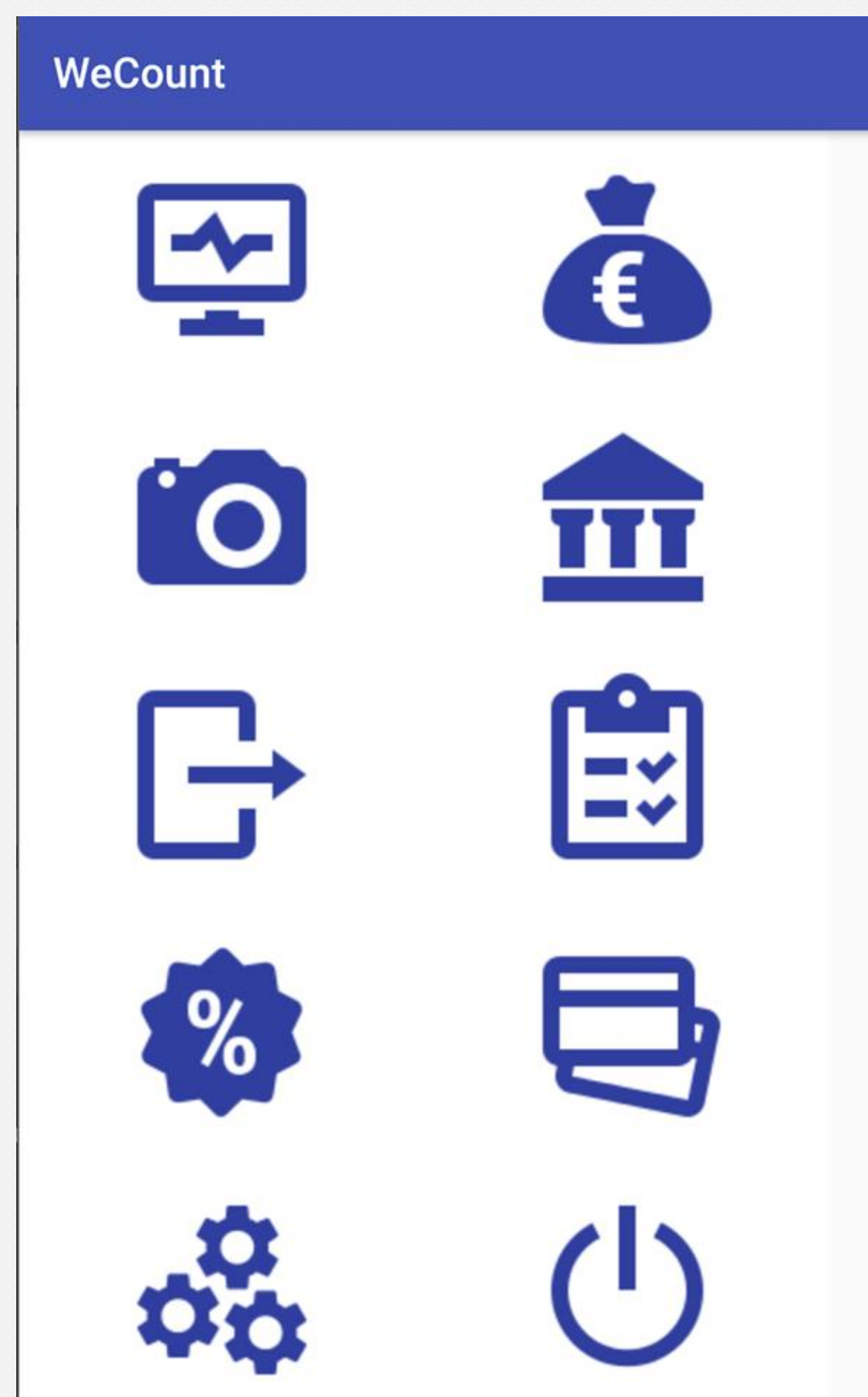
- Réalisation d'entretiens semi-directifs sur différentes classes d'âges. Intérêt dénoté pour le suivi de budget, l'alerte à l'approche d'un certain pallier de dépense et la liste de course prédictive.

IA

- Analyse du contenu du ticket de caisse (parsing suivi d'un KNN sur 2 couches, opérant selon la distance de Levenshtein entre le texte détecté et les noms des produits normalisés de la BDD, puis leurs abréviations).
- Analyse des habitudes d'achat de l'utilisateur afin de lui proposer une liste de course pertinente, où les articles sont ordonnés par priorité (travail sur la fréquence d'achat, la date de dernier achat...).

CLIENT SERVEUR

- Assurer l'intégralité des communications entre le client et le serveur. Utilisation de Socket sécurisées. Images et objets complexes sont échangées au format JSON.



Quelques éléments bibliographiques

- Comprendre la distance de Levenshtein : https://fr.wikipedia.org/wiki/Distance_de_Levenshtein
 Comprendre le principe du KNN : https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm#Metric_learning
 Elimination de Gauss Jordan : https://fr.wikipedia.org/wiki/%C3%89limination_de_Gauss-Jordan
 Grafikart.fr, comprendre les bases de données : <https://www.youtube.com/watch?v=qrmLVJBJ-D4&list=PLjwdMgw5TTLUJLpzUYGBK7K5-hPgZA7zo>
 Crawler en java pour récupérer une page html : <https://github.com/yasserg/crawler4j>
 Documentation pour parser une page html : <https://jsoup.org/http://www.lagis.univ-lille1.fr/~bonnet/image/OpGeo.pdf>
 « Ils ont innové dans un monde numérique »
www.Yves-Poillane@telecom-paristech.fr