

DÉTECTION D'INTENTION : APPLICATION INDUSTRIELLE D'UN PROJET DE RECHERCHE

APPLICATION À L'INTENTION D'ACHAT

Estelle Maudet – Ingénieure de Recherche

Julien Noblat – Ingénieur Développement

Daniel Dubois – Responsable Marketing

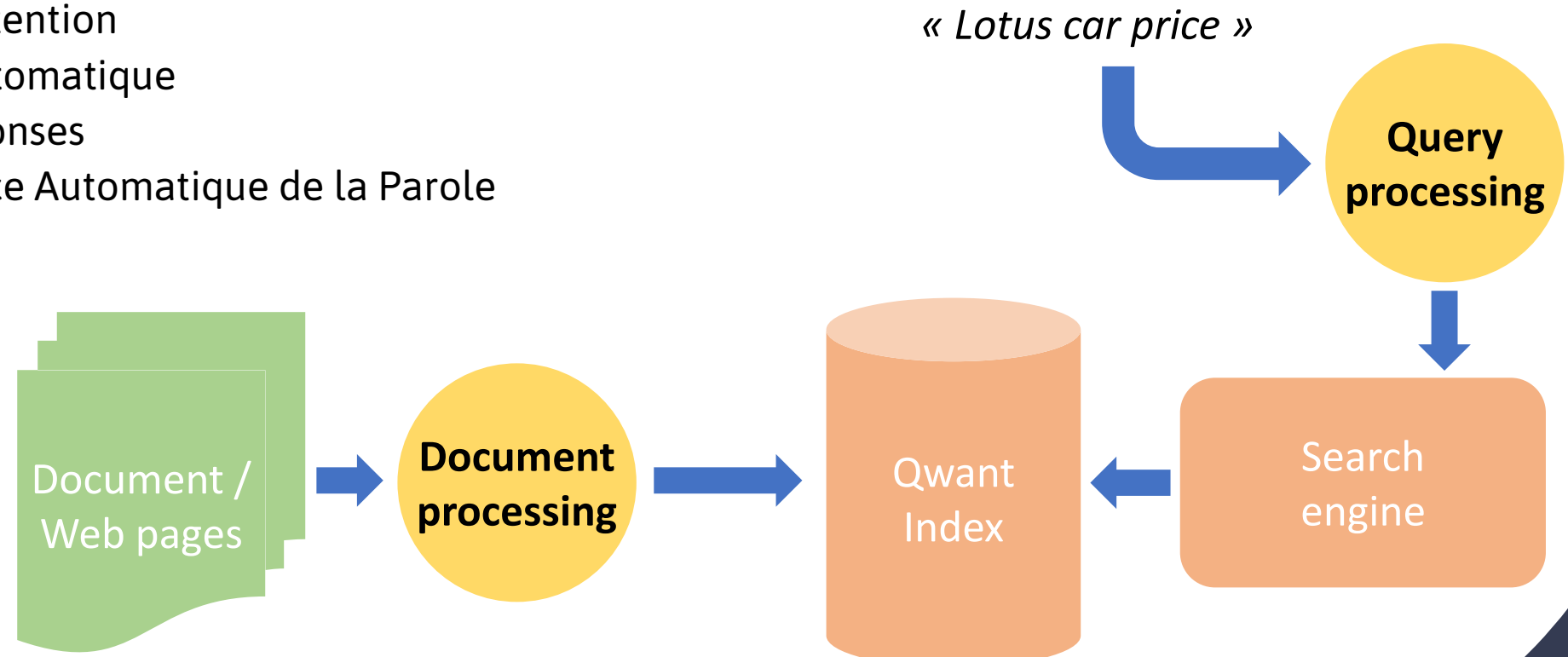
Christophe SERVAN, PhD – Responsable Scientifique



Qwant

TRAITEMENT AUTOMATIQUE DES LANGUES

- Le **TAL** et le **Deep Learning** sont essentiels pour un moteur de recherche tel que Qwant
- Intégration de dans le moteur de recherche à travers le traitement des **documents et des requêtes** :
 - Reconnaissance d'Entités Nommées (REN) comme les dates, localisations, noms, etc.
 - Compréhension de la langue naturelle (NLU) la partie sémantique
 - Détection d'intention
 - Traduction Automatique
 - Question/Réponses
 - Reconnaissance Automatique de la Parole



**ETUDE D'UN CAS
PRATIQUE :**

**LA DETECTION
D'INTENTION
(SHOPPING)**



DÉTECTION D'INTENTION

- Traitement au niveau de la requête
- Tâche de classification :
 - Informationnelle (une donnée en particulier)
 - Date de naissance ?
 - Navigationnelle (un site web en particulier)
 - wikipedia.fr
 - Transactionnelle (une action particulière)
 - Shopping
 - ...



DÉTECTION D'INTENTION

- Traitement au niveau de la requête
- Tâche de classification :
 - Informationnelle (une donnée en particulier)
 - Date de naissance ?
 - Navigationnelle (un site web en particulier)
 - wikipedia.fr
 - Transactionnelle (une action particulière)
 - **Shopping**
 - ...



DÉTECTION D'INTENTION: CLASSIFICATION LÉGÈRE

Contraintes:

- Qualité des modèles
 - Précision
 - Rappel
 - F-mesure
- Performance des modèles
 - Temps de traitement



DÉTECTION D'INTENTION: CLASSIFICATION LÉGÈRE

Contraintes:

- Qualité des modèles
 - Précision
 - Rappel
 - F-mesure
- Performance des modèles
 - Temps de traitement
- **Pas de données annotées**



**CRÉATION DE DONNÉES
&
CYCLES D'APPRENTISSAGE**

DÉTECTION D'INTENTION: ABSENCE DE DONNÉES ANNOTÉES

Données:

- Absence de données annotées
- Présence de données non-structurées:
 - Requêtes
 - Logs
 - Liste de produits



DÉTECTION D'INTENTION: ABSENCE DE DONNÉES ANNOTÉES

Données:

- Absence de données annotées
- Présence de données non-structurées:
 - Requêtes
 - Logs
 - Liste de produits
- **Créer des données annotées**



APPRENTISSAGE FAIBLEMENT SUPERVISÉ

- Extraction automatique d'informations
 - Requêtes & clic logs, bases de connaissances (liste de produits), etc.
- Génération automatique d'exemple à partir de patrons (templates)
 - e.g.: Restaurant \${CITY}
- Clustering
 - Extraction automatique de partition
 - Sélection manuelle de partition

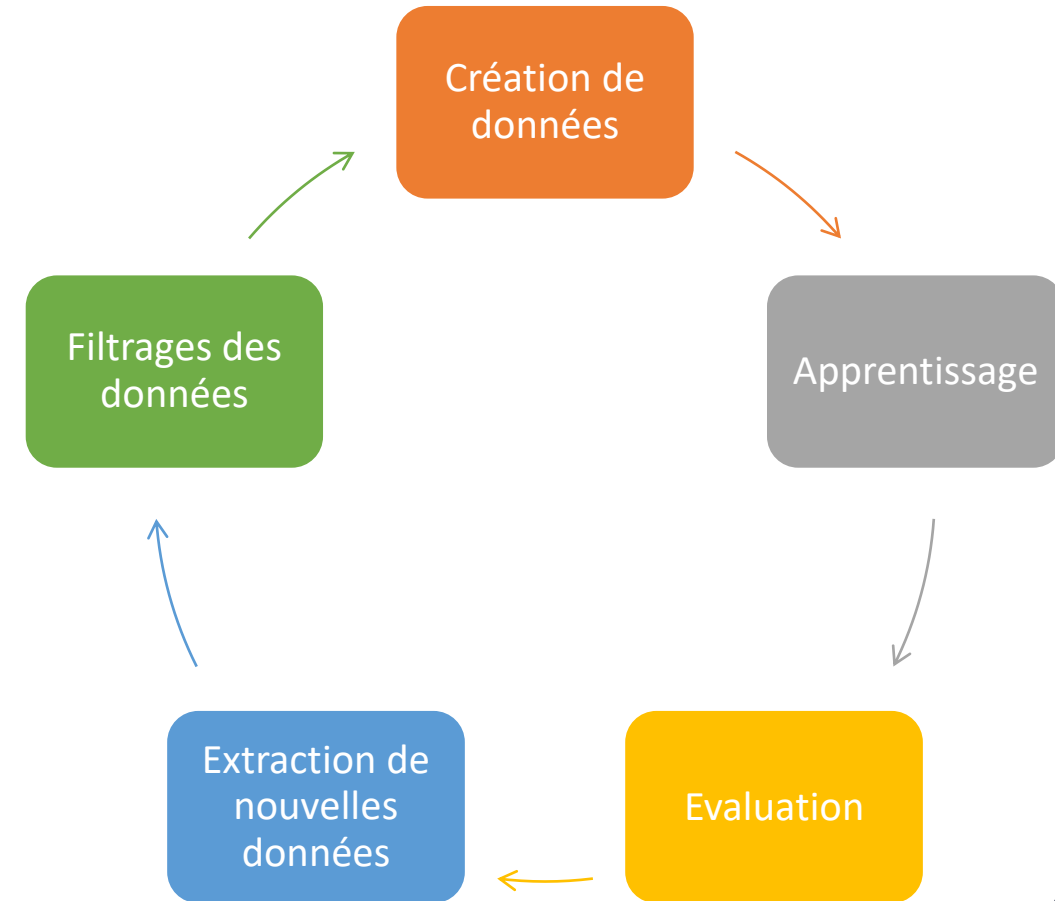
APPRENTISSAGE FAIBLEMENT SUPERVISÉ

- Extraction automatique d'informations
 - Requêtes & clic logs, bases de connaissances (liste de produits), etc.
- Génération automatique d'exemple à partir de patrons (templates)
 - e.g.: Restaurant \${CITY}
- Clustering
 - Extraction automatique de partition
 - Sélection manuelle de partition

⇒ **Human in the loop**

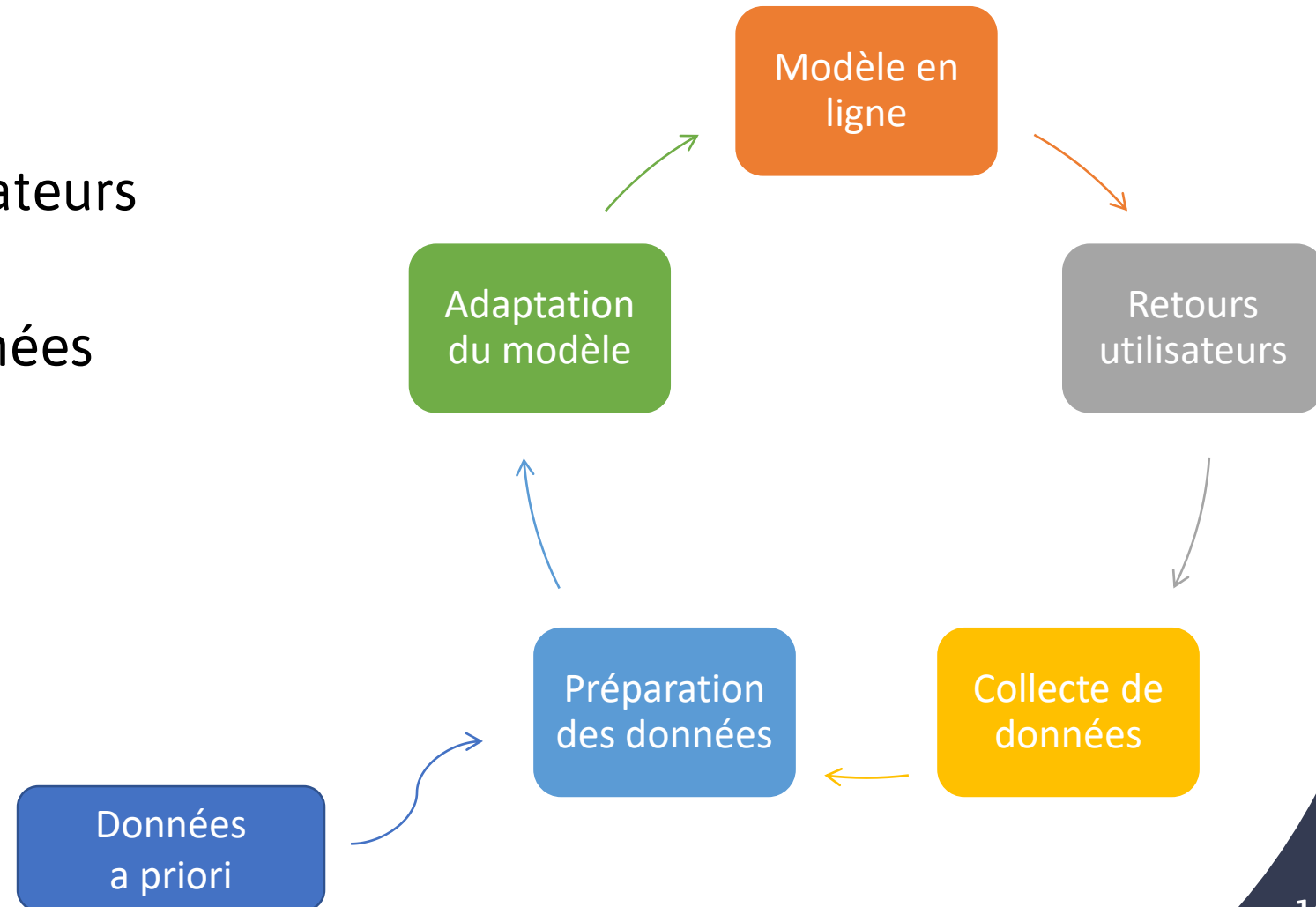
MODÈLE V0 : BOOTSTRAPPING

- A partir d'une faible quantité de données
- Apprentissage et évaluation (vérification)
- Appliquer le nouveau modèle (nouvelles données extraites)
- Filtrage / Sélection / corrections
- Création d'un nouveau corpus d'apprentissage



MODÈLE V1 : APPRENTISSAGE CONTINU

- Modèles en production
- Extraction des retours utilisateurs
- Collecte et filtrage des données
- Préparation des données (Automatiques et a priori)
- Adaptation des modèles



APPROCHE DE CLASSIFICATION

DÉTECTION D'INTENTION: CLASSIFICATION LÉGÈRE

Principales approches explorées:

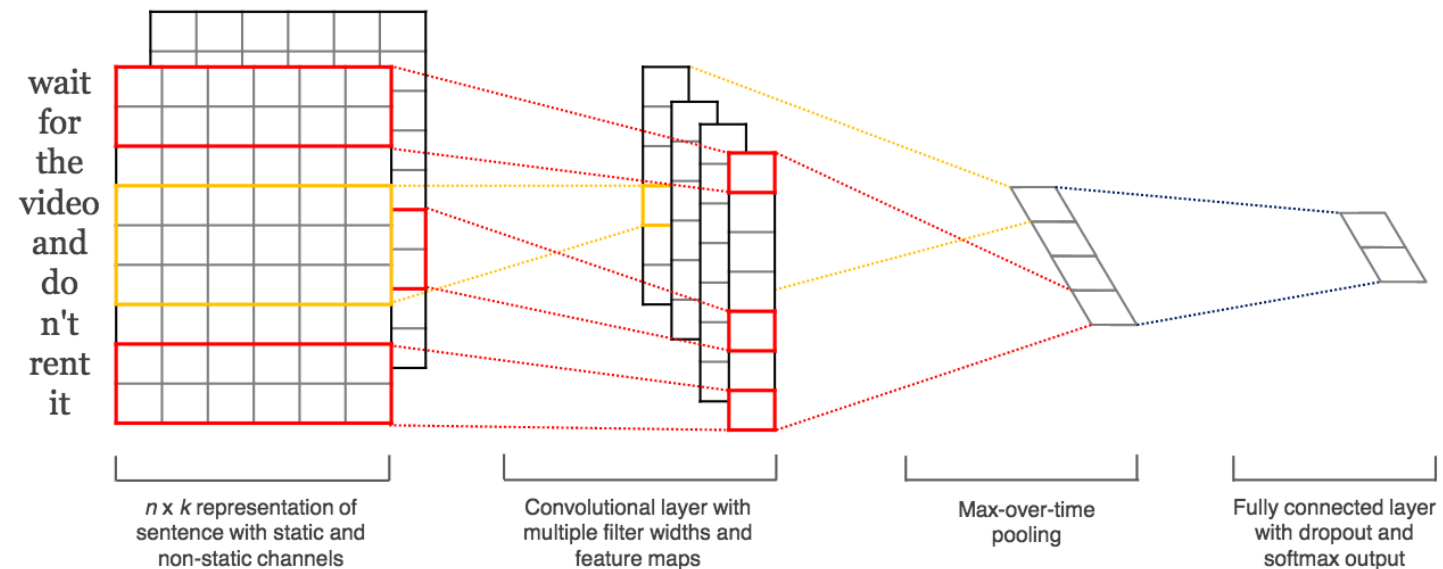
- Matrice de convolution sur les représentations de mots + max pooling + softmax [Kim, 2014]



DÉTECTION D'INTENTION: CLASSIFICATION LÉGÈRE

Principales approches explorées:

- Matrice de convolution sur les représentations de mots + max pooling + softmax [Kim, 2014]



DÉTECTION D'INTENTION: CLASSIFICATION LÉGÈRE

Principales approches explorées:

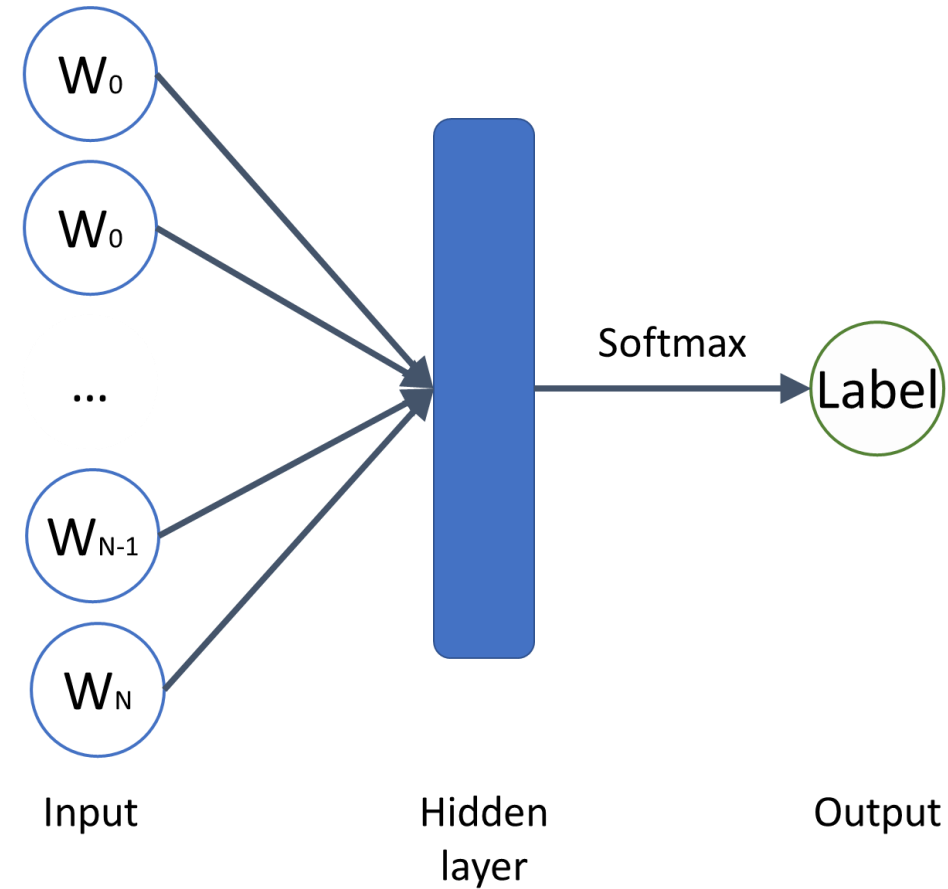
- Matrice de convolution sur les représentations de mots + max pooling + softmax [Kim, 2014]
- Choix « the simplest the better » : Une couche simple + h-softmax + dropout



DÉTECTION D'INTENTION: CLASSIFICATION LÉGÈRE

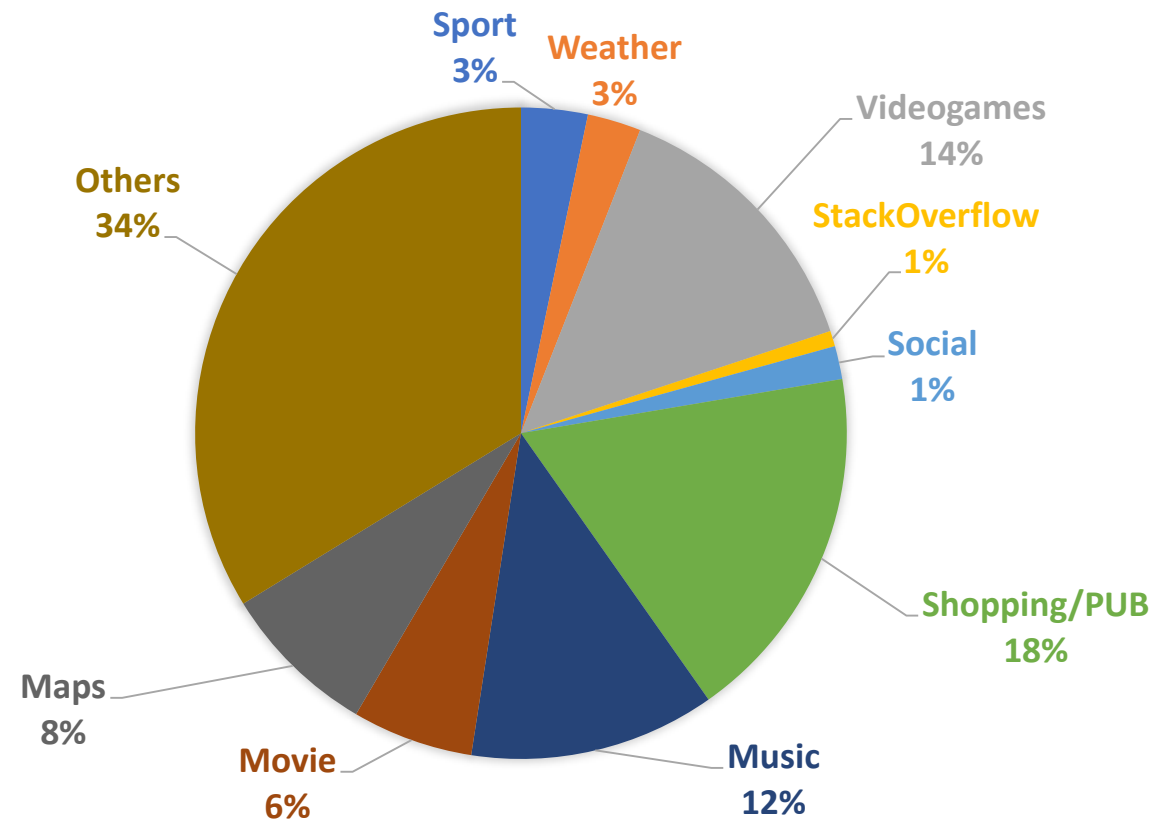
Principales approches explorées:

- Matrice de convolution sur les représentations de mots + max pooling + softmax [Kim, 2014]
- Choix « the simplest the better » : Une couche simple + h-softmax + dropout



DONNÉES DE DÉPART

TRAINING DATA REPARTITION



EVALUATION

Corpus de test annoté manuellement (10 intentions)

Modèle	F-mesure
V0 – Proof of Concept	85,70

EVALUATION

Corpus de test annoté manuellement (shopping)

Modèle	F-mesure
V0 – Proof of Concept	44,40
V1 – Pré-production	60,20
V2 – A/B Test	89,20
V3 – final	90,50

EVALUATION

Corpus de test annoté manuellement (shopping)

Modèle	F-mesure (10 intentions)	F-mesure (shopping)
V0 – Proof of Concept	85,70	44,40
V1 – Pré-production	85,50	60,20
V2 – A/B Test	83,30	89,20
V3 – final	83,60	90,50

MISE EN OEUVRE TECHNIQUE

DÉTECTION D'INTENTION: OBJECTIFS DE PERFORMANCES

API:

- Méthode de communication
 - API REST
- Structure des données
 - Json
- Etude d'implémentations
 - Python
 - C++



DÉTECTION D'INTENTION: OBJECTIFS DE PERFORMANCES

Python :

- Code rapidement
- Compétences courantes
- Modules existants
- Traitements lents

C++ :

- Code plus lentement
- Compétences en C++ plus rares
- Pas ou peu de modules existants
- Traitements très rapides



DÉTECTION D'INTENTION: OBJECTIFS DE PERFORMANCES

Python :

- Code rapidement
- Compétences courantes
- Modules existants
- **Traitements lents**

C++ :

- Code plus lentement
- Compétences en C++ plus rares
- Pas ou peu de modules existants
- **Traitements très rapides**



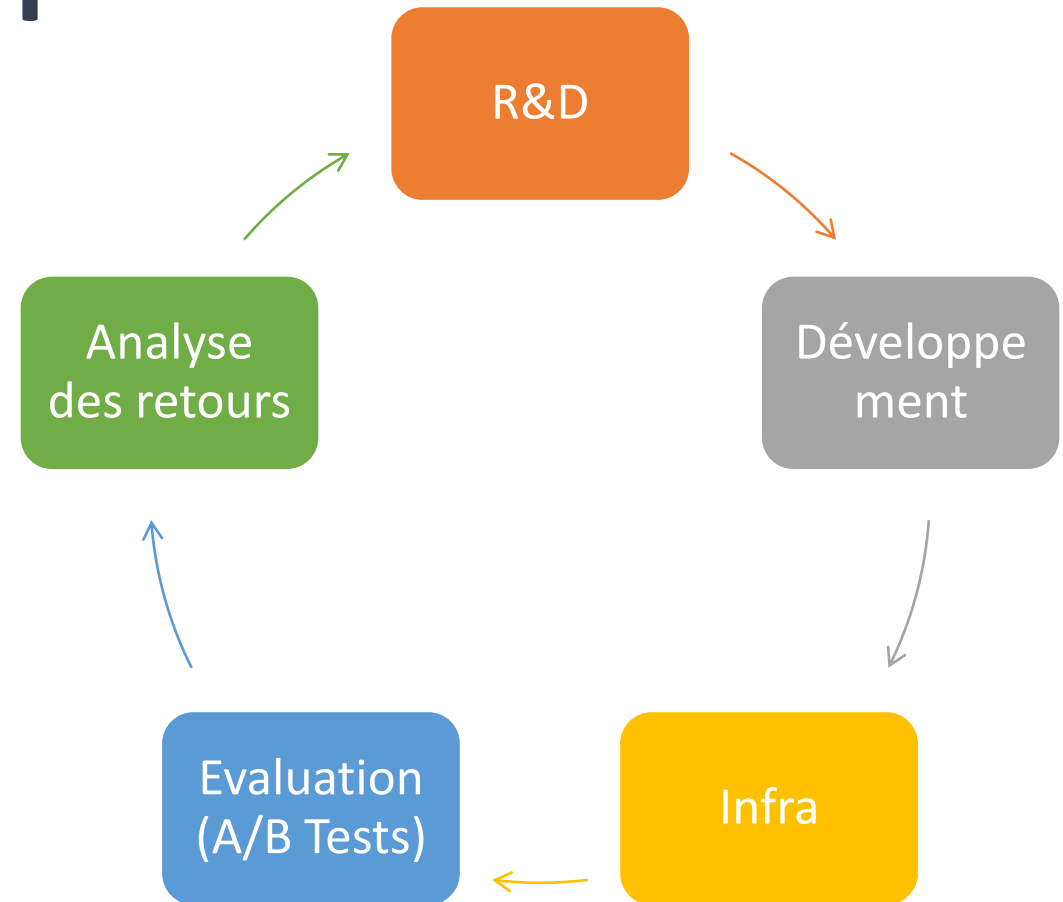
DÉTECTION D'INTENTION: DÉVELOPPEMENT DE L'API

Cycle :

- Etablissement d'un cahier des charges
- Coordination de 3 services internes
- Evaluations périodiques

Dev en C++ :

- Alpha : en 1 mois (pré-production)
- Beta : en 3 mois (A/B tests)
- Release V0 : 2 mois (production)



EVALUATION DES PERFORMANCES

Utilisation de Apache Benchmark

Modèle	Temps de réponse moyen	Nbr requêtes / sec
V0 – Proof of Concept – Python	22ms	45,5
V1 – Beta – C++	0,45ms	≈ 2 222

CONCLUSION



DETECTION D'INTENTION

- Exploration de contraintes réelles
 - Pas ou peu de données
 - Performances
 - Qualité
- Utilisation SoA
 - Réseaux de neurones
- Exploitation des possibilités
 - C++

DETECTION D'INTENTION

- Exploration de contraintes réelles
 - Pas ou peu de données
 - Performances
 - Qualité
- Utilisation SoA
 - Réseaux de neurones
- Exploitation des possibilités
 - C++
- **Mise en production avec succès (+30% de trafic & de revenus)**

MERCI DE VOTRE ATTENTION !
QUESTIONS ?

[c.servan\[AT\]qwant\[DOT\]com](mailto:c.servan@qwant.com)

