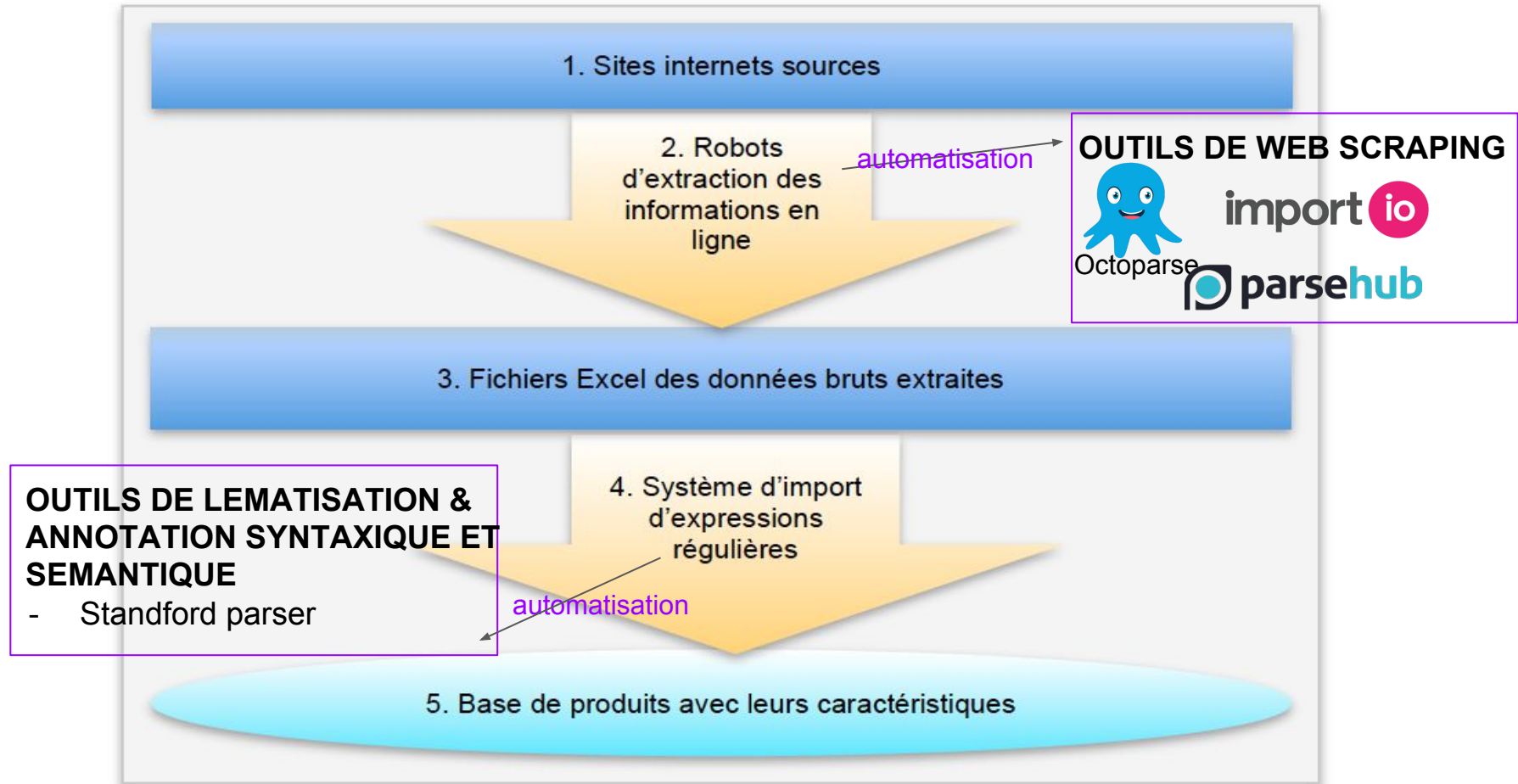


1. Modélisation du domaine : création de l'ontologie

- actuellement manuelle (recherche sur le domaine + modélisation dans l'outil CMS)
- possibilité d'automatisation : GATE
 - à partir de textes bruts...
 - création des types de caractéristiques (couleur)
 - créer des caractéristiques (rouge, bleu ...)
 - défaut:
 - pas capable de peupler avec des instance (Montre connecté Pebble : couleur=rouge)
 - recherches en cours...



2. Peuplement de la base



Octoparse



Pourquoi data?



Fournir des renseignements opportuns aux utilisateurs finaux



Surveiller les données Web en temps réel



Analyser les tendances du marché



Enrichissement le contenu de produit (Conseiller digital, contenu pour Wordpress...)

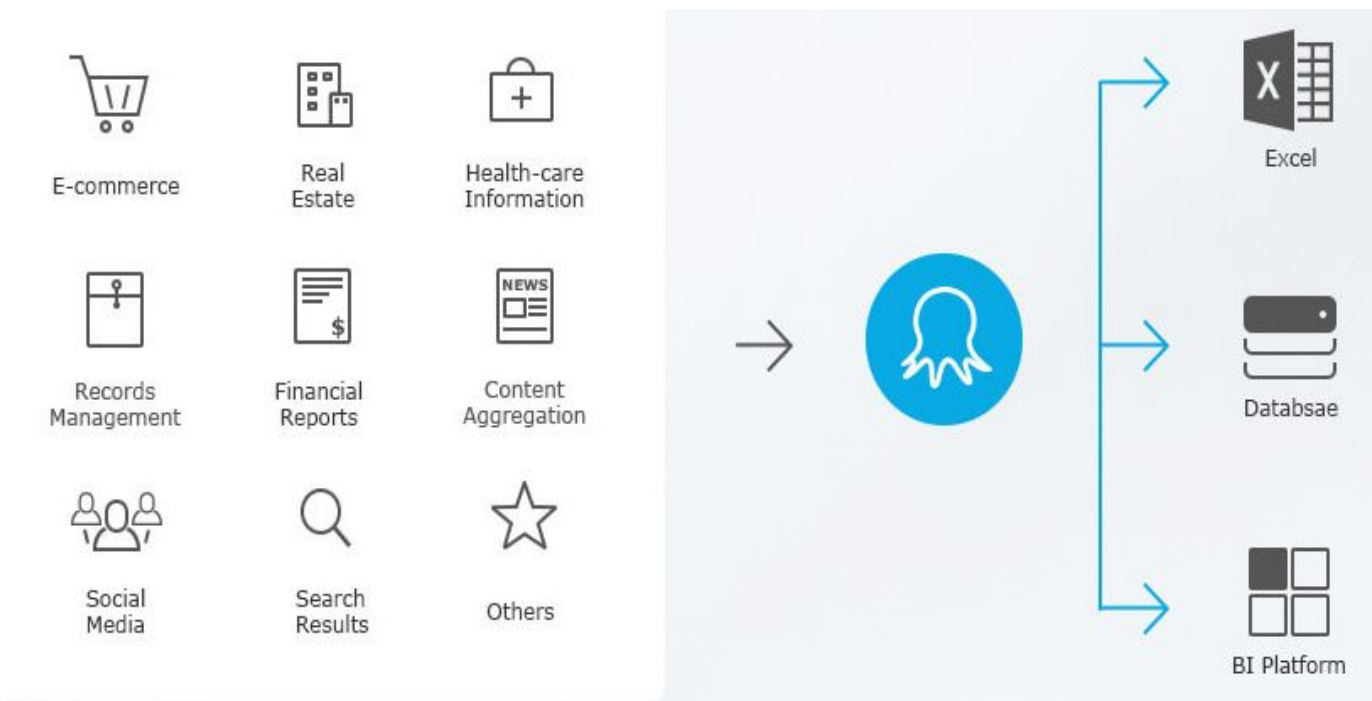
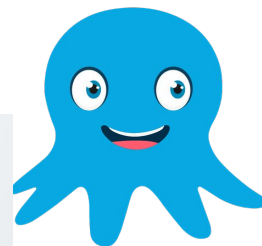


L'analyser des sentiments de client (Commentaires et critiques sur nos FB, blog, produits...)



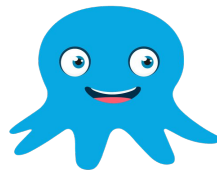
Rassemblez veille concurrentielle

Comment?



<http://www.octoparse.com/>

Points forts d'Octoparse



Simple: Click Interface

Utiliser un chemin de Xpath et des expressions régulières pour recueillir les données avec une précision

Toutes sortes de données en cours de chargement

Crawler les données de toutes sortes de techniques de chargement de données telles que AJAX, ou JavaScript.

Service cloud

Soutien Proxy et API, rotation IP automatique pour empêcher IP d'être bloqué

Serveur Cloud

Un scraping rapide

Reconnaissance de l'image

L'interface OCR permet de crawler les textes sur les images.

Fonctions automatiques

Demarrage automatique à scraping à l'heure régulière.

Gratuit

Tous les fonctions.

Autres utilisation des outils de web crawling à des fins marketing

1. Obtenir des liens de qualités vers son site (LinkBuilding)

TROUVER DES LIENS
CASSES OU VERS DES
SOURCES QUI
N'EXISTENT PLUS

- Broken Link Finder
- Check my Links
- Link Prospector

OUTILS DE
RECHERCHE DE
BACKLINKS



Mais aussi ...

- analyse le positionnement des concurrents (mots clefs...)

OUTILS POUR
CONTACTER DES
PROPRIÉTAIRES DE SITES



Mais aussi ...

- trouve les contacts influents dans 1 domaine

Utilisation des outils de web crawling à des fins marketing

2. Trouver des contacts influents et de blogueurs dans un domaine



3. Trouver des idées de mots clefs dans un domaine ou par rapport à un texte



4. Optimiser du contenu

- nTopic : scanne un contenu et donne des suggestion pour le rendre plus SEO-friendly
- MarketMuse : scanne un contenu et le compare à d'autres contenus similaires sur le web pour suggérer d'y inclure certains termes clés

Le web sémantique

Web 1.0

- début du web (1995-2000)
- pages webs reliées par des liens hypertextes
- information static

Web 2.0

- réseaux sociaux, blogs ...
- interactif : écrire des commentaires, mettre à jours son profil

Web 3.0

- le web sémantique ?
- favorise le partage des données entre les différentes applis
- structuration de la connaissance pour qu'elle soit compréhensible par la machine (RDF, OWL, XML)
- modélisation du monde (Google Knowledge Graph)

Sources

<http://backlinko.com/high-quality-backlinks> (LinkBuilding)

<http://www.inkybee.com/>

<http://www.scrapebox.com/>

<http://www.webmarketing-conseil.fr/outils-seo/#outil-positions-google>

<http://www.referralsaasquatch.com/how-to-use-web-scraping-for-marketing/>