

# Deloitte.

# **CRAWLER FOR TMAS PLATFORM**

**OCTOBER 2019**

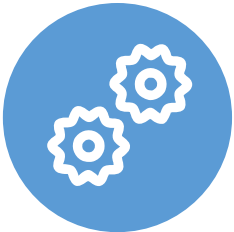
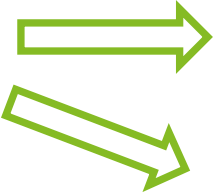


# General Description

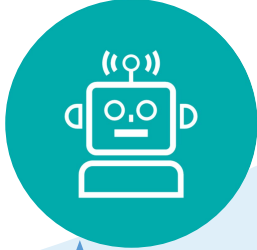


**Crawler - is a part of RoBMS solution.**  
**RoBMS - Robotic Benchmarking Study platform.**

**Artificial intelligence for benchmarking**



50+ man-hours



20% less



Current development stage

80% less

### Manual Process

- **Obligatory for tax purposes**
- **Judgmental and subjective decision**
- **Sub-optimal cost**
- **Hundreds of studies per year worldwide**

### Partial solution / Web Crawler

- **Automated search of companies' web-sites or other sources with activity description**

### Complete solution

- **Activity description based on Natural language recognition and processing**
- **Automated comparison of companies' activities**
- **Better accuracy of results**

# RoBMS Overview

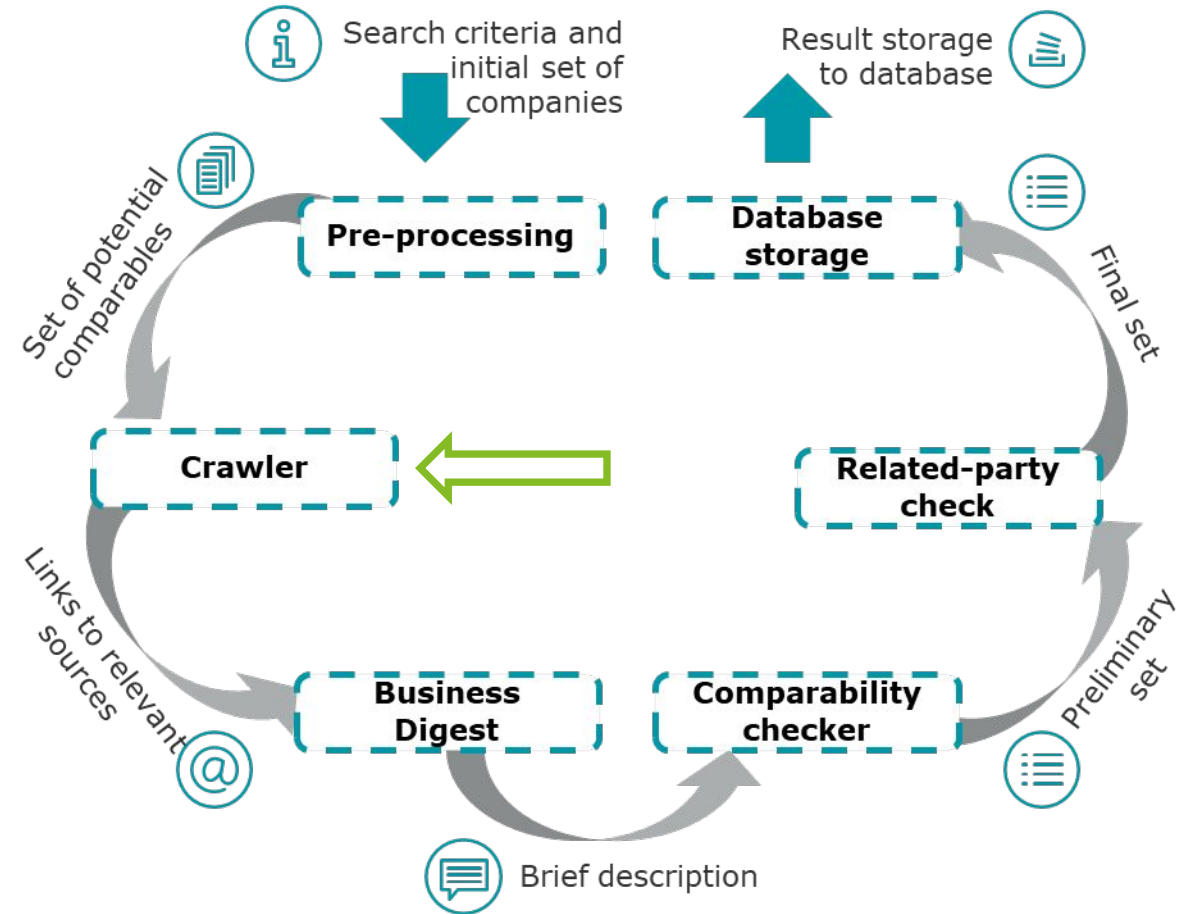


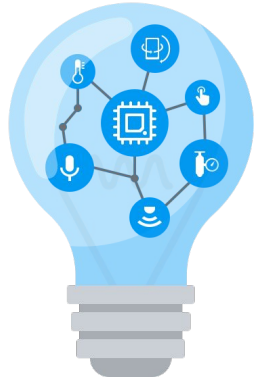
## Problem

**Transfer pricing worldwide projects require performance of benchmarking study, and considered to be the manual time-consuming task, which diverts highly-skilled resources for trivial tasks.**

## Solution

- **Enabling to reduce study time consumption up to 80%**
- **Preprocessing list of potentially comparable companies from external database**
- **Building list of websites for study**
- **Robotic company web-site scraping**
- **Text analytics to build structured activity digest for all companies in analysis**
- **Automated decisions on comparable companies based on cognitive analysis**
- **Related/Non-related party decision based on text analytics and external database info**
- **Results storage**





## Machine Learning Text Analysis Algorithms

- **Neural networks based on Natural Language Processing (NLP) algorithms**
- **Relevant text detection**
- **Improved machine translation algorithms (based on Google API service)**
- **Text comparison**

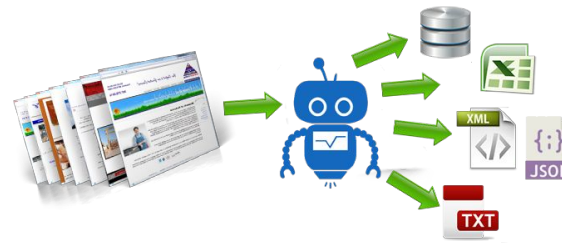
## Web Crawling Data Mining

- **Rules-based web-search for each company;**
- **Website relevance check;**
- **Text extraction from web-sites.**



## Data Storage & Management

- **Message Query management**
- **Translation results caching and storage**
- **Web-sites caching and storage**
- **Final results storage**



## CIM INSIGHTS

- Provides smart outputs like the estimated market size, key industry trends and future outlook based on the repository of CIMs (anonymized)
- Provides company specific financial trends by leveraging the web crawler that periodically runs through
- **Enhances** understanding of insights received, aging of the various responses, etc. and trends around the industry/company (how perception has changed, market share fluctuations, etc.), including future outlook
- **Feedback loop: Input to Target Screening app for drawing intelligent insights**



## COMPENSATION BENCHMARKING

- Leverage web crawlers to scrub public information sources to enhance the benchmarking database (Glassdoor, etc.)
- Utilize cognitive insights to provide additional insights into new-age attributes that should be considered in compensation benchmarking such as company maturity/growth, business model etc.
- Provide insights into the latest industry-wide trends in compensation benchmarks



## SYNERGY VALUE DRIVERS

- **Add a web crawler that scrubs for announced synergy and realization timelines in various publicly announced deals**
- **Provides recommended value capture ideas based on input provided by the user on current deal / engagement, by utilizing the intelligence acquired based on past experiences.**
- **Utilize cognitive intelligence to provide additional insights on synergy realization timeline, key risks and mitigation plan, etc.**



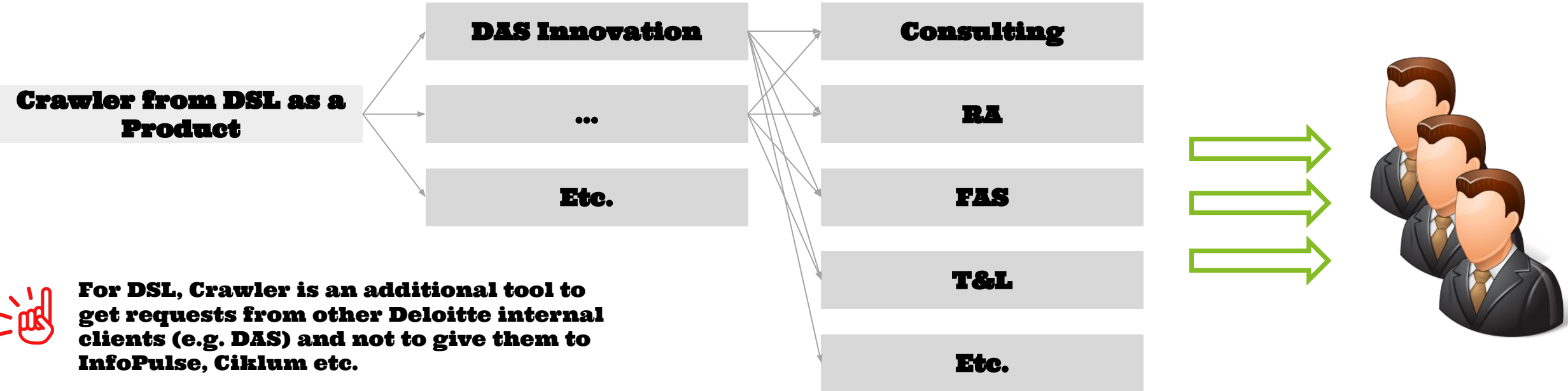
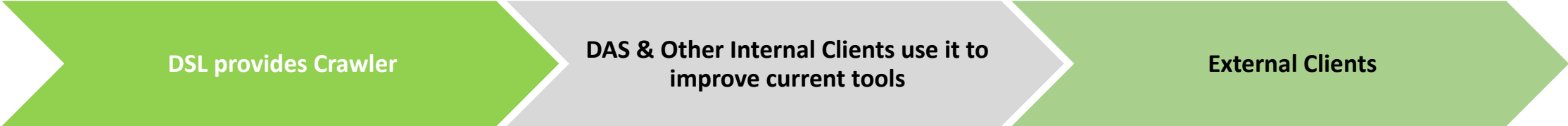
## SYNERGY AND ONE-TIME COST BENCHMARKING

- **Integrates with the Total M&A Solution and Thrive tools to perform analytics on the synergy and one-time cost benchmarks**
- **Links to the relevant deliverables should practitioner want to deep dive into a particular benchmark figure**
- **Leverages / Integrates with a web crawler that periodically runs through public sources to track synergy announcements and improve the database**





# Client's Scheme



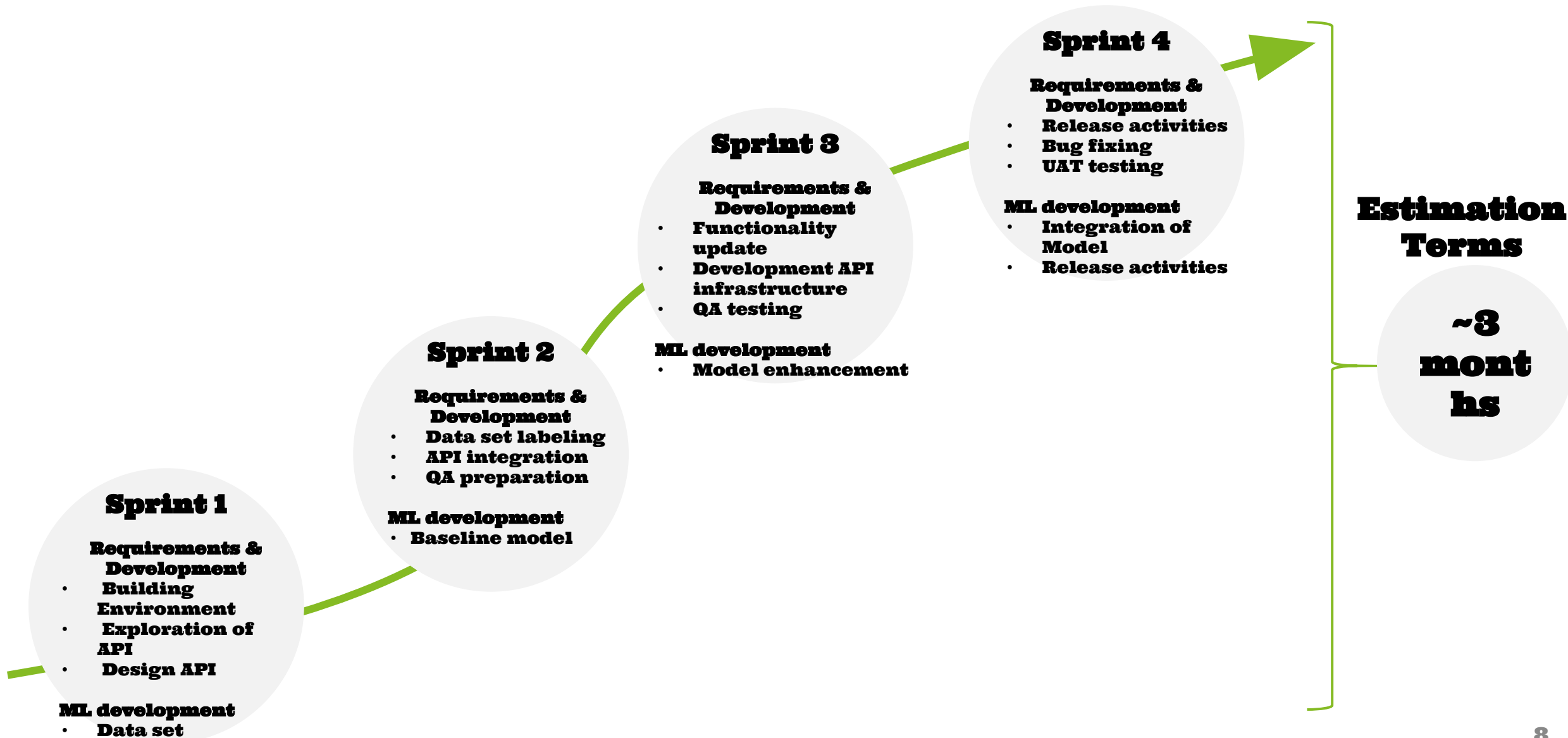
**For DSL, Crawler is an additional tool to get requests from other Deloitte internal clients (e.g. DAS) and not to give them to InfoPulse, Ciklum etc.**



**For business-units Crawler is an additional tool to provide External clients more services and to create additional revenue stream.**

# Possible Work Plan

D.





# Team & Costs

D.

Resources	Rates, \$/Hour	Sprint 1	Sprint 2	Sprint 3	Sprint 4
Senior Python Developer / Tech Lead	38	40	40	40	40
Middle Python Developer	29	40	40	40	40
Business analyst/Product owner	34	20	20	20	20
NLP Specialist Senior	38	80	80	80	80
NLP Specialist Middle	29	80	80	80	80
QA Manual Middle	24	20	60	80	80
DevOps CI/CD Senior	34	40	10	10	20
Data Labeling Specialist	15		200		

Hours per Sprint	320	550	370	320
Costs Per Sprint	\$10,560	\$13,500	\$10,980	\$11,320
<b>TOTAL</b>	<b>\$ 46,360</b>			



- **Technical lead - 1**
- **Product Owner - 1**
- **ML Developer - 2**
- **Python Developer - 2**
- **QA manual - 1**
- **DevOps - 1**



- ✓ **Resources effort, costs = \$ 46,360**
- ✓ **Contingency = \$ 4,636**
- ✓ **Total = \$ 70,966**

# Delivery Team & Baseline Solution



## **Crawler Baseline**

**version = \$ 20,000**

### **Crawler baseline solution includes:**

- **Crawling web, using keywords and client profile**
- **Ranking algorithm for more relevant results**
- **ML model that highlights most important information on the web pages**
- **Dockerized solution with simple maintenance API**

**SPECIAL OFFER**

## **UA Delivery team**

- **Building Environment**
- **Deployment existing baseline solution into environment**
- **Updating structure of existing API**
- **Data set preparation and labeling**
- **API integration**
- **Model development**
- **Functionality update**
- **Development of API infrastructure**
- **QA testing**

# Crawler's Advantages

---

# D.

## Key benefits of implementation



**Optimal targeting**



**Opportunity and risks management**



**Cost saving**





# Deloitte.

# THANK YOU



#### **About Deloitte**

**Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee (“DTTL”), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as “Deloitte Global”) does not provide services to clients. Please see [www.deloitte.com/about](http://www.deloitte.com/about) for a detailed description of DTTL and its member firms. Please see [www.deloitte.com/us/about](http://www.deloitte.com/us/about) for a detailed description of the legal structure of Deloitte LLP and its subsidiaries. Certain services may not be available to attest clients under the rules and regulations of public accounting.**

**Copyright © 2017 Deloitte Development LLC. All rights reserved.  
Member of Deloitte Touche Tohmatsu Limited**

