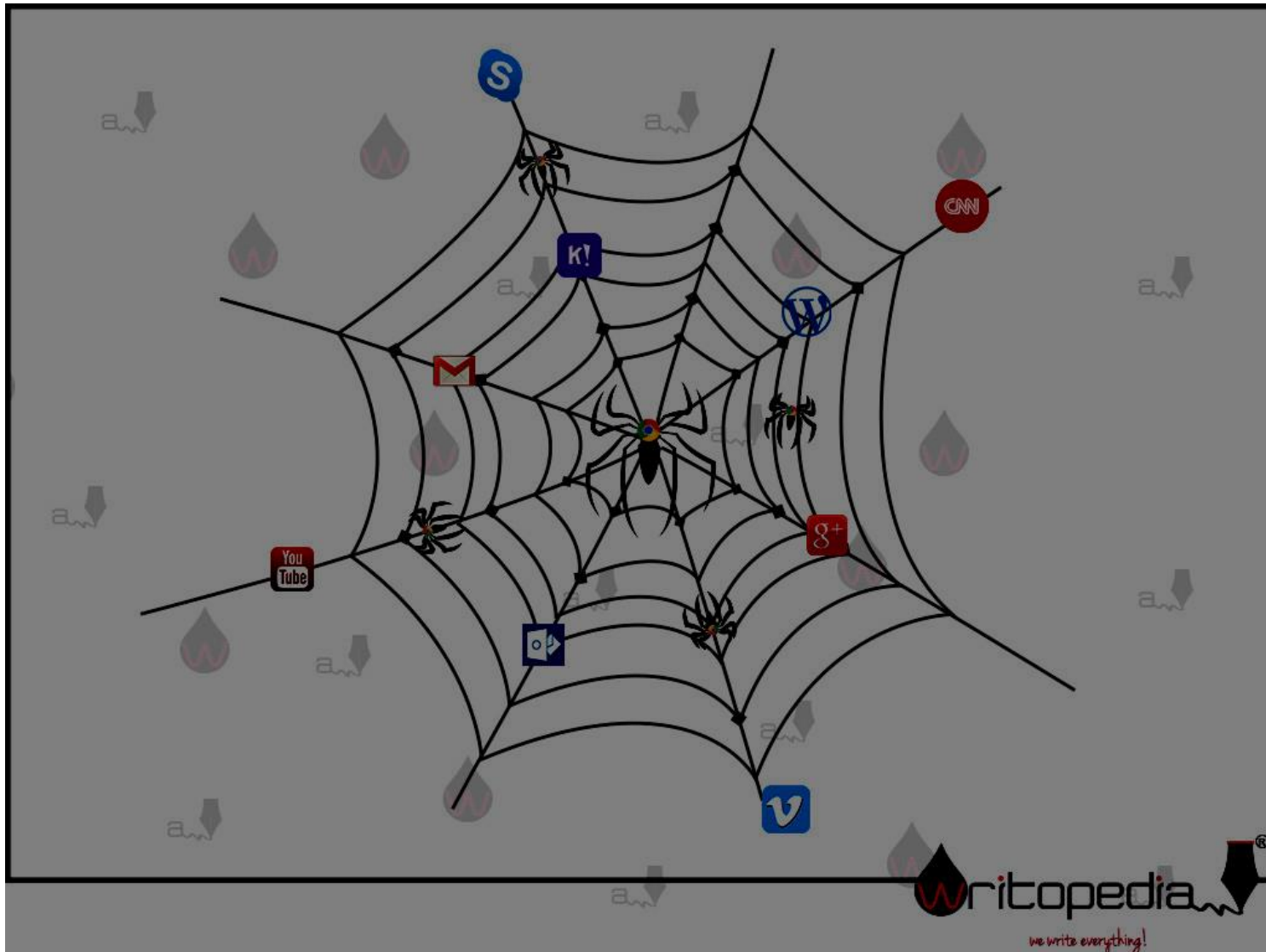


WEB CRAWLING



SCRAPY

.PYTHON

.CÓDIGO ABERTO

.INTERFACE SIMPLES PARA CONSTRUÇÃO DE SPIDERS

.FACILIDADE DE ESTRUTURAR CONTEÚDO EXTRAÍDO

.EXECUÇÃO DE SPIDERS NA NUVEM(CLOUD)

PRIMEIROS PASSOS

.INSTALAR O SCRAPY:

```
~/Codigos/scrapy$ pip install scrapy
```

.CRIAR UM NOVO PROJETO:

```
~/Codigos/scrapy$ scrapy startproject teste
```

ESTRUTURA DO PROJETO

```
└─ teste
    └─ spiders
        └─ __init__.py
        └─ __init__.py
        └─ items.py
        └─ middlewares.py
        └─ pipelines.py
        └─ settings.py
    └─ scrapy.cfg
```

MANIPULANDO HTML

html

```
<div class="list-cities">  
  <ul>  
    <li>  
      <a href="/cidades/1">Gotham</a>  
    </li>  
  </ul>  
</div>
```

xpath

```
//div[@class="list-cities"]//ul//li//a
```