**SEARCH ENGINES!!!!**

One side of application crawls stack overflow and other error message sites + enable own crawler that is able to crawl documentation sites in general using key words:

* <https://www.youtube.com/watch?v=AFigoBtWNQ0>
* Google custom search engine returns a large amount of web pages. Filter through internally and decide on most applicable result based on accuracy to key word searched?
* Figure out results and work backwards

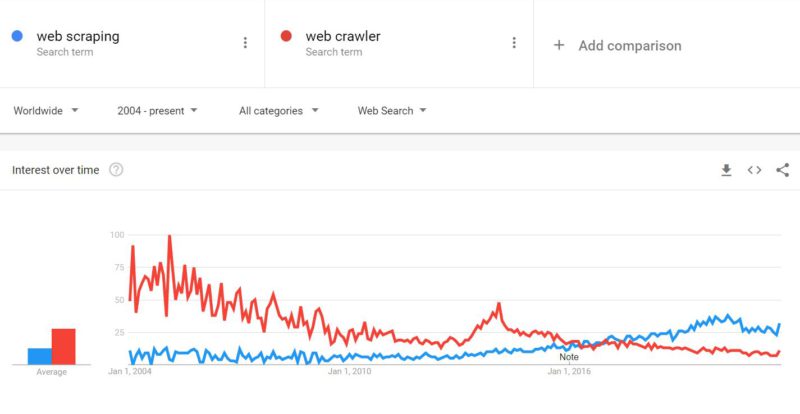
<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.896.2014&rep=rep1&type=pdf>

1. Option for results?:

Chart

Description automatically generated

Define 10 Documentation sites to crawl and perform a similar analysis on the similarity score - <https://www.researchgate.net/profile/Ashish-Kumar-Maurya/publication/276206236_URL_Ordering_based_Performance_Evaluation_of_Web_Crawler/links/555c088e08ae8f66f3ade28e/URL-Ordering-based-Performance-Evaluation-of-Web-Crawler.pdf>



Rise of scrapers over crawlers example - <https://research.aimultiple.com/web-crawler/>

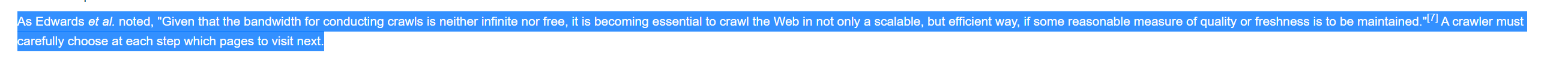
<https://research.aimultiple.com/web-crawler/>

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

COPY AND PASTE JOB (mention in web scrapers):

According to Google data, interest in the web crawler topic has decreased since 2004. Yet, at the same time period, interest in web scraping has outpaced the interest in web crawling. Various interpretations can be made,some are:

* Increasing interest in [analytics](https://research.aimultiple.com/analytics/) and [data-driven decision making](https://research.aimultiple.com/data-driven-decision-making/) are the main drivers for companies to invest in scraping.
* Crawling done by search engines is no longer a topic of increasing interest since they have done this since the early 2000s
* Search engine industry is a mature industry dominated by Google and Baidu, so few companies need to build crawlers.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

^ Reword

Graphical user interface, text

Description automatically generated

<https://www.researchgate.net/profile/Filippo-Ricca/publication/220135806_Web_crawlers_compared/links/54894a710cf268d28f090a14/Web-crawlers-compared.pdf>

Pattern recognition algorithms : <https://programmerall.com/article/3245289882/>

Qualities of a good web crawler: <https://www.promptcloud.com/blog/important-qualities-of-good-web-crawler/>