

**IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF TEXAS
TYLER DIVISION**

SPIDER SEARCH ANALYTICS LLC

Plaintiff,

v.

HUMAN DATA PROJECT, INC.,

Defendant.

§
§
§
§
§
§
§
§
§
§

CIVIL ACTION NO.

JURY TRIAL DEMANDED

COMPLAINT FOR INFRINGEMENT OF PATENT

COMES NOW, Plaintiff Spider Search Analytics LLC (“SSA” or Plaintiff), through the undersigned attorneys, and respectfully alleges, states, and prays as follows:

NATURE OF THE ACTION

1. This is an action for patent infringement under the Patent Laws of the United States, Title 35 United States Code (“U.S.C.”) to prevent and enjoin defendant Human Data Project, Inc. d/b/a HDP Health, (hereinafter “Defendant”) from infringing and profiting, in an illegal and unauthorized manner and without authorization and/or of the consent from SSA, from U.S. Patent No. 7,454,430 (the “430 patent”, attached hereto as Exhibit “A”) (the “Patent-in-Suit”) pursuant to 35 U.S.C. § 271, and to recover damages, attorney’s fees, and costs.

THE PARTIES

2. Plaintiff SSA is a Texas entity with its principal place of business at 101 E. Park Blvd, Suite 600, Plano, Texas 75074.

3. Upon information and belief, Defendant is a company organized and existing under the laws of the State of Delaware, having a principal place of business at 680 Mission St., San

Francisco, California 94105. Upon information and belief, Defendant may be served with process through its agent for service of process The Company Corporation, at 2711 Centerville RD Ste 400, Wilmington, DE 19808.

JURISDICTION AND VENUE

4. The Court has subject matter jurisdiction over this action pursuant to 28 U.S.C. §§ 1331 and 1338(a) because the action arises under the Patent Laws of the United States, 35 U.S.C. §§ 1 *et seq.*

5. Defendant is subject to this Court's personal jurisdiction pursuant to due process and/or the Texas Long Arm Statute, due at least to its substantial business and purposeful availment of this forum, including: (i) at least a portion of the infringements alleged herein; and (ii) regularly doing or soliciting business, engaging in other persistent courses of conduct, and/or deriving substantial revenue from goods and services provided to individuals in Texas and in this judicial district.

6. Upon information and belief, Defendant, directly and/or through its employees or agents, and/or its customers, uses products, as defined below, that perform each and every step of at least one claim of the '430 patent with the knowledge and/or understanding that such products are used or will be used in this District. For example, Defendant uses the accused product to extract data from web pages.¹ Upon information and belief, Defendant has engaged in substantial and not isolated activity within this District. Therefore, exercise of jurisdiction over Defendant will not offend traditional notions of fair play and substantial justice. Such an exercise is consistent with the Texas long-arm statute.

¹ <http://www.apifier.com>.

7. Venue is proper in this judicial district pursuant to 28 U.S.C. §§ 1391 and 1400(b) because Defendant is subject to personal jurisdiction in this district, has regularly conducted business in this judicial district and certain of the acts complained of herein occurred in this judicial district.

FACTUAL ALLEGATIONS

8. On November 18, 2008, the United States Patent and Trademark Office (“USPTO”) duly and legally issued the ‘430 patent, entitled “System and method for facts extraction and domain knowledge repository creation from unstructured and semi-structured documents” after a full and fair examination. (Exhibit A).

9. SSA is presently the owner of the patent, having received all right, title and interest in and to the ‘430 patent from the previous assignee of record. SSA possesses all rights of recovery under the ‘430 patent, including the exclusive right to recover for past infringement.

10. The ‘430 patent contains three (3) independent claims and twenty-four (24) dependent claims.

11. The ‘430 patent claims, *inter alia*, a method for building a deep web crawler.

12. Defendant uses, *inter alia*, a method that performs each and every step of at least one claim of the ‘430 patent.

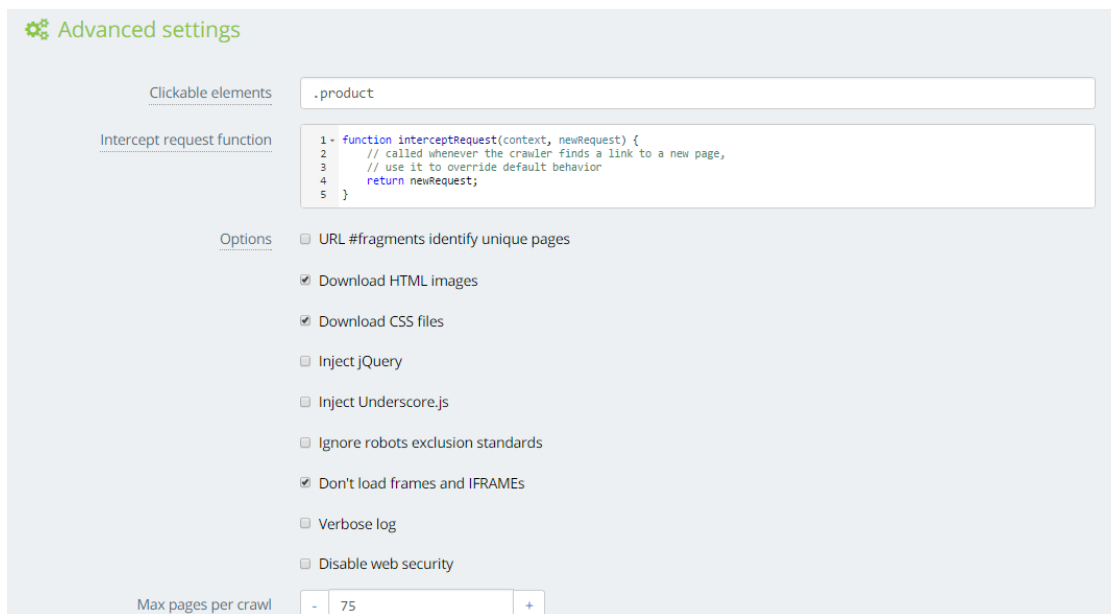
DEFENDANT’S PRODUCTS

13. In accordance with claim 10 of the ‘430 patent Defendant uses a method for building a deep web crawler. For example, Defendant is a client of the Apifier company and uses the “Apifier web scraper” (the “Accused Product”) in order to extract structured data from websites.² The Accused Product allows customers to build web crawlers that are capable of

² <https://www.apifier.com/customers>.

crawling any website (including, upon information and belief, websites from the “deep web” and that use Javascript and AJAX for dynamic pages).

14. In accordance with claim 10 of the ‘430 patent Defendant utilizes scout crawling rules to collect dynamic pages. For example, Apifier’s customers specify a number of scout crawling rules such as: Crawl pseudo-URLs, Clickable Elements, Max Pages Per Crawl, Max Crawl Depth, Execution Timeout, Resource Timeout, Page Load Timeout, Max Pages Per IP Address, and Max Parallel Processes.



The image shows a configuration interface for a crawler. It consists of ten rows, each with a label on the left and a control on the right. The controls are either text input fields with minus and plus buttons, or dropdown menus with unit labels. The settings are as follows:

Setting	Value	Unit
Max result records	[Empty]	
Max crawling depth	5	
Execution timeout	[Empty]	s
Resource timeout	[Empty]	ms
Page load timeout	[Empty]	ms
Page function timeout	[Empty]	ms
Infinite scroll height	[Empty]	px
Delay between requests	2000	ms
Max pages per IP address	[Empty]	
Parallel crawling processes	1	

Source: Example UI for creating a Crawler using Apifier's website.

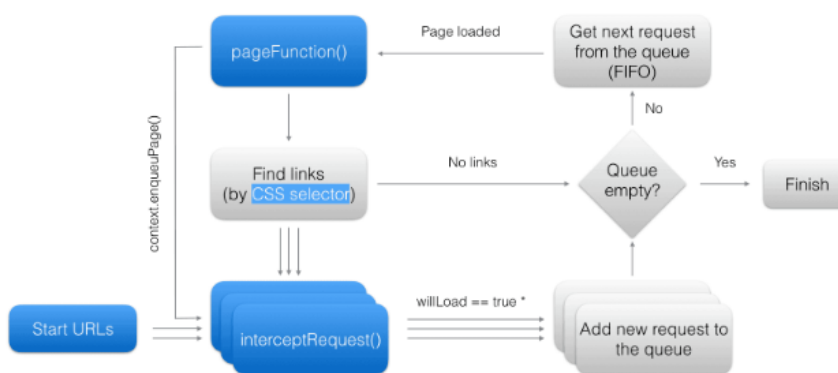
15. In accordance with claim 10 of the '430 patent Defendant utilizes an analyzer and extractor to determine underlying structure of queries. For example, Defendant provides a list of Start URLs which are added into the crawl queue. Apifier's analyzer then utilizes a CSS selector (Clickable Elements) to analyze the HTML elements on crawled pages to determine the underlying structure of queries, including but not limited to the links on those pages. The analyzer uses extractor code to extract the HTML and identify hyperlinks for this purpose. The links so identified are further analyzed by the analyzer to determine which links meet the user-specified Crawl Pseudo URL criteria. The qualifying links are then fed into the crawl queue.

Crawling process

First, let's discuss the crawling process of Apifier. By default, the crawler repeats the following steps:

1. Add each of the *Start URLs* into the crawling queue
2. Fetch the first URL from the queue and load it in the virtual browser
3. Execute *Page function* on the loaded page and save its results
4. Find all links from the page. If a link matches any of the *Crawl pseudo-URLs* and has not yet been enqueued, add it to the queue
5. Go to step 2.

This process is depicted in the diagram below. Note that blue elements represent settings or operations that can be modified in crawler settings.



Source: <http://support.apifier.com/knowledgebase/articles/863403-crawling-multiple-pages>.

To enable the crawling of more pages, we need to define *Crawl pseudo-URLs* (also called PURLs). PURL is simply a URL with JavaScript-style regular expressions enclosed in `[]` brackets. This PURL will be matched against every link URL when evaluating whether the URL should be enqueued.

If you click the *More* link at `https://news.ycombinator.com/news`, you'll be sent to `https://news.ycombinator.com/news?p=2`. The next click on that link will navigate to `https://news.ycombinator.com/news?p=3`, and so on. A PURL matching these pages is simply `"https://news.ycombinator.com/news?p=[\d+]"`. Note that the regular expression we used `(\d+)` means one or more digits. To learn more about regular expressions, check [this W3Schools tutorial](#) or [regexp101.com](#).

When we add the above PURL to our crawler configuration, the crawler should theoretically start loading the next pages. However, when you run the crawler you will see that it still finishes after the first page. The problem lies in the *Clickable elements* options in *Advanced settings*, because for some reason the default setting skips the *More* link.

Source: <http://support.apifier.com/knowledgebase/articles/863403-crawling-multiple-pages>.

16. In accordance with claim 10 of the '430 patent Defendant generates instructions for a harvester, wherein the harvester provides requests to a server and collects available pages from

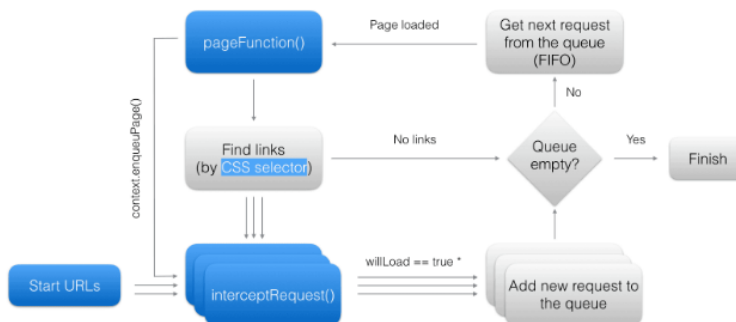
the server. For example, once the analyzer adds all the Start URLs and the relevant links that were found on the pages corresponding to the Start URLs, to the crawl queue, Defendant generates the instructions for a harvester and the harvester begins transmitting requests to the server for the web pages represented by the entries in the crawl queue.

Crawling process

First, let's discuss the crawling process of Apifier. By default, the crawler repeats the following steps:

1. Add each of the *Start URLs* into the crawling queue
2. Fetch the first URL from the queue and load it in the virtual browser
3. Execute *Page function* on the loaded page and save its results
4. Find all links from the page. If a link matches any of the *Crawl pseudo-URLs* and has not yet been enqueued, add it to the queue
5. Go to step 2.

This process is depicted in the diagram below. Note that blue elements represent settings or operations that can be modified in crawler settings.



17. The elements described in paragraphs 13-16 are covered by at least claim 10 of the '430 patent.

INFRINGEMENT OF THE '430 PATENT

18. Plaintiff realleges and incorporates by reference the allegations set forth in paragraphs 1 to 17.

19. In violation of 35 U.S.C. § 271, Defendant is now, and has been directly infringing the '430 patent.

20. Defendant has had knowledge of infringement of the '430 patent at least as of the service of the present complaint.

21. Defendant has directly infringed and continues to directly infringe at least claim 10 of the '430 patent by using the Accused Product without authority in the United States, and will continue to do so unless enjoined by this Court. For example, Defendant's website, which is available in Texas, utilizes and implements the Accused Product and performs each and every step recited in claim 10 of the '430 patent.³ As a direct and proximate result of Defendant's direct infringement of the '430 patent, Plaintiff has been and continues to be damaged.

22. By engaging in the conduct described herein, Defendant has injured SSA and is thus liable for infringement of the '430 patent, pursuant to 35 U.S.C. § 271.

23. Defendant has committed these acts of infringement without license or authorization.

24. As a result of Defendant's infringement of the '430 patent, SSA has suffered monetary damages and is entitled to a monetary judgment in an amount adequate to compensate for Defendant's past infringement, together with interests and costs.

25. SSA will continue to suffer damages in the future unless Defendant's infringing activities are enjoined by this Court. As such, SSA is entitled to compensation for any continuing and/or future infringement up until the date that Defendant is finally and permanently enjoined from further infringement.

DEMAND FOR JURY TRIAL

26. SSA demands a trial by jury of any and all causes of action.

PRAYER FOR RELIEF

³ *Id.*

WHEREFORE, SSA prays for the following relief:

1. That Defendant be adjudged to have infringed the Patent-in-Suit directly, literally and/or under the doctrine of equivalents;
2. That Defendant, its officers, directors, agents, servants, employees, attorneys, affiliates, divisions, branches, parents, and those persons in active concert or participation with any of them, be permanently restrained and enjoined from directly infringing the Patent-in-Suit;
3. An award of damages pursuant to 35 U.S.C. §284 sufficient to compensate SSA for the Defendant's past infringement and any continuing or future infringement up until the date that Defendant is finally and permanently enjoined from further infringement, including compensatory damages;
4. An assessment of pre-judgment and post-judgment interest and costs against Defendant, together with an award of such interest and costs, in accordance with 35 U.S.C. §284;
5. That Defendant be directed to pay enhanced damages, including SSA's attorneys' fees incurred in connection with this lawsuit pursuant to 35 U.S.C. §285; and
6. That SSA have such other and further relief as this Court may deem just and proper.

Dated: April 6, 2017

Respectfully Submitted,

By: /s/Eugenio J. Torres-Oyola

Eugenio J. Torres-Oyola

USDC No. 215505

Ferraiuoli LLC

221 Plaza, 5th Floor

221 Ponce de León Avenue

San Juan, PR 00917

Telephone: (787) 766-7000

Facsimile: (787) 766-7001

Email: etorres@ferraiuoli.com

Jean G. Vidal Font

USDC No. 227811

Ferraiuoli LLC

221 Plaza, 5th Floor

221 Ponce de León Avenue

San Juan, PR 00917

Telephone: (787) 766-7000

Facsimile: (787) 766-7001

Email: jvidal@ferraiuoli.com

**ATTORNEYS FOR PLAINTIFF
SPIDER SEARCH ANALYTICS LLC**