

How Employbl Saved 250 Hours Building Their Career-Matching Database

Fixing tech starts with hiring. And fixing hiring is an information problem. That's what Connor Leach, cofounder and CEO at Employbl discovered when creating a new talent marketplace meant to connect tech employees with the information-rich hiring marketplace they deserve.

Tech job seekers rely on a range of metrics to gauge the opportunity and stability of a potential employer.

While information like funding rounds, founders, team size, industry, and investors are often public, it can be hard to grab the myriad fields candidates value in a up-to-date format from around the web.

These difficulties are amplified by the fact that many tech startups are often "long tail" entities that also regularly change.

To solve these issues, Employbl partnered with Diffbot to gain web-wide firmographic coverage for brand new and established orgs.



and

EMPLOYBL

Key Takeaways

- Though a great deal of firmographic data is public, manual and rule-based extraction is hard to scale
- Rich insights in firmographic data are a function of how up-to-date and comprehensive data sources can be.
- Even in a single -- if broad -- field like tech, can require web data extraction from a TON of domains.

The Problem

Employbl started with a good deal of information. Their database contained info meant for job seekers on over 1,000 tech companies. The issue was that there was no central location to update this data from. If they continued manually, valuable time for an early stage startup would be eaten up. Their index likely wouldn't get larger as too much time would be spent just keeping it up to date. If they continued scraping, the range of sites they needed information from would lead to a ton of scraper maintenance (and each scraper would need to be tailored to each site).

In short, Employbl needed web-wide coverage for large and small tech companies that updated when company sites updated. And they needed this data in a structured format with useful fields users could filter companies with.

Employbl originally tried Clearbit for enriching organizational data. But found that there was no service tier aligned with newly started orgs or solo entrepreneurs. Additionally, Clearbit would charge the same amount whether an organizational data enrichment call returned data or not.

Products Used



Diffbot's Knowledge Graph is the world's largest knowledge graph sourced from public web data. Cutting-edge AI-enabled web extraction APIs are passed data from crawled pages to turn the unstructured web into structured queryable knowledge. Billions of interlinked entities including organizations, news mentions, and people help to power use cases like Employbl's.

Extraction Challenges

Many data teams spend up to 80% of their time and resources aggregating and cleansing data that may be messy, unstructured, outdated, or downright inaccurate.

How Employbl Built Out a Richer Database in Less Time

Manually, it was taking Employbl about 3 minutes to update each company in their database. At 5,000 companies, that would be over 250 hours of time! Alternatively, setting up scraping of their own may have taken weeks or months and would have required maintenance.

Diffbot's Knowledge Graph and corresponding API provide a wide-ranging ontology of relationships between -- and facts about -- hundreds of millions of organizations and people.

Data in the Knowledge Graph is sourced from the public web, with an average of over 20 "facts" per entity, and an average of 6 "origins" for each fact. This allows for greater data coverage and precision in a range of fields, with no manual extraction or custom build scrapers.

Knowledge Graph data was able to provide Employbl up-to-date data from across the web including information on funding rounds, investors, number of employees, location, detailed industry data, and more. All details individual seeking to work in tech find important when job searching!

What This Meant For Employbl

The major roadblock to Employbl's focus on democratizing tech and startup hiring data was access to scalable data. By handing off the task of crawling and structuring the public web, Employbl was able to focus on more meaningful developments to their platform.

A Structured Web

Diffbot's Automatic Extraction APIs, Knowledge Graph, and Enhance products provide access to the largest collection of web data commercially available.

Our AI-enabled extraction and processing systems parse the web into contextually linked entities including organizations, products, people, articles, and more.

While clients can point automatic extraction APIs to sites of their choosing, Diffbot also crawls the entirety of the web and parses it into semantic data every 3-5 days.

Currently this data is available in the Knowledge Graph in the form of over 10 billion entities and 3 trillion facts.

In addition to being able to expand their database by thousands of companies, Diffbot's structured web data was able to expand their range of fields attached to firmographic data as well as provide data provenance.

Throughout the process over 250 hours of time was saved when compared to their prior method! That's a ton of time for an early stage startup and allowed their team to move on to matters more important than data gathering.

Public Web Data For The Public

Employbl doesn't charge for their platform. And a primary driver in the creation of the platform included helping those affected by Covid-19 related layoffs in Silicon Valley. The conspicuous lack of enterprise data enrichment tools at a price point of a new enterprise or individual entrepreneurs is telling. With many data enrichment platforms you're paying for shadily bought data, rather than harnessing what's already public on the web. Diffbot structures "guilt-free" public web data so you can query the web like a database and provide public data back to the public in a more usable format.

Firmographic Data Toolkit

Data on hundreds of millions of companies.

Linked data including top management, employees, skills, news mentions.

Robust filtering by over 20 fields (industry codes, location, subsidiaries, founding date, employees, market cap, descriptions, many others).

Graphable relationships.

Access within our Knowledge Graph dashboard as well as integrations in Excel, Google Sheets, and Zapier.

Want to supercharge your firmographic coverage?
Start a Free Trial Today or contact us at sales@diffbot.com