



Intelligent Data Extraction Automation (IDEA) Search Tool

Building Databases through a Single Click

Contents

Executive Summary..... 3

Introduction 3

Challenges 3

Solution 3

Nexii Solution 4

Business Benefits 5

Summary 5

Introduction

Data collection from professional networking site accounts consumes significant amounts of time. The user has to manually collect the data from the portal save it in an organized manner, and validate it before displaying the valid data. To overcome the above hurdles, Nexiilabs developed IDEA (Intelligent Data Extraction through Automation) a search automation tool that collects, manages, and validates the data with minimal manual interaction and efforts.

Executive summary

Building databases is a critical function for businesses. Always in demand these lists are the go-to resource for reasons ranging from recruitment and staffing to business development, procurement and networking purposes. Extracting the right data can however be an uphill task given the vastness of the data pool. Manual searches to identify, extract, and build a database list can be a laborious and time consuming exercise. Having identified this gap, Nexii Labs developed a tool that automates the search function helping businesses optimize their searches to generate best results in least time.

In this paper, we discuss the Intelligent Data Extraction Automation (IDEA) search tool developed by Nexii IT Labs.

Challenges

Manual searches have certain limitations. For example – A person can collect data of maximum **100** profiles per day. However he cannot validate the data to ensure the email address collected is correct or not. Search for the organizational email address of the person demands significantly more time. At the same there is also the possibility of duplicate data being collected.

Solution

Nexiilabs designed an automation solution that can extract data in a fraction of the time usually taken, and validate before saving it in an organized manner.

Nexiilabs Solution

IDEA is a solution for the problem defined. We opted Selenium with Java and designed a process to collect the data from the networking account after opening it in a browser. A Graphical User Interface is developed for the user convenience to place a new request, to view the existing requests, data of the processed request, configuring the mail servers and mail boxes, etc. It is a Java application included the usage of frameworks like springs, Hibernate, etc.

The tool collects the requests from the users with a few mandatory fields like Country, Industry, Designation, number of contacts required. Then the tool navigates to the professional networking site account and collects the data by visiting the filtered data on it. The data collected with fields like Name, Designation, Country, Industry, etc. are saved in the database, validates the email domain of that particular company. If the email domain is valid, it creates 7 different formats of email addresses and are saved in the database.

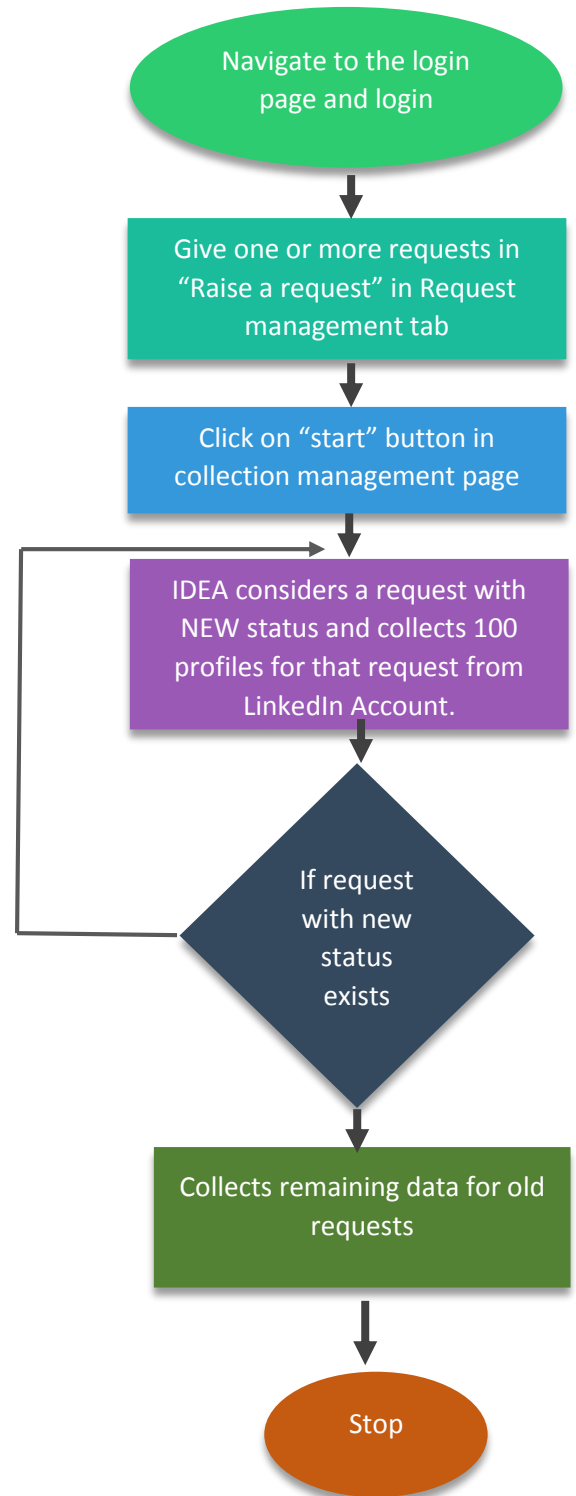


Fig 1. IDEA tool process

An **Email Validator** which is also a part of the IDEA, sends emails to the **7** formats of the email addresses and check for the bounces. It filters the invalid email addresses and displays the data with valid email address for particular request.

The tool collects the data continuously processing one request after the other.

Business benefits

- User's data collecting capability rises to an average of 250 - 300 or more contacts per day as compared to just a maximum of 100 contacts.
- Email validation becomes very easy with IDEA email Validator.
- IDEA reduces the data collection and list build time
- Data accuracy is optimized.
- Data duplication is minimized.

Summary

The IDEA tool supports businesses by enabling easy collection of precise data in the shortest possible time, through a single click of a mouse.



Nexii Labs is a US headquartered IT solutions provider with an R&D center in Hyderabad, India. With deep domain expertise in expertise in storage, virtualization and cloud technologies, we offer state-of-the-art application development, quality assurance, mobility solutions, cloud services, managed services and technical support services. Our clients list includes global leaders and Fortune 500 businesses. We are also an ISO 27001 and ISO 9001:2008 certified organization.

To know more, visit us on: www.nexiilabs.com

Mail us at: info@nexiilabs.com

Follow us at:   