

DOI: <https://doi.org/10.24297/ijct.v23i.9559>**Experimental Implementation approach of Crawling Framework CARE JOBS for Disability Job Search Using Intelligent Agents**

Diana Avram

Grafen Software, Romania

diana.avram@grafensoft.com

Abstract

The Crawling Framework CAREJOBS (CF) we present here is a suite of tools dedicated to facilitating job search for people with disabilities. In this experimental implementation we propose a crawler with modern architecture based on intelligent agents, a multi-agent system trained on the basis of a rule engine. We also present a case of using CF in a CV creation expert. It can be seen that our approach is perfectible by adding learning to the agent model.

Keywords: Intelligent agents, Crawling Framework, Disability, Job Search

I. Introduction

We are currently witnessing a significant real-time digital transformation of our society, resulting in an expanding "digital economy" that greatly influences the entire job market. Alongside the opportunities arising from this transformation, there are challenges, such as the need to ensure that no one is left behind in the digitization process, including over one billion individuals with disabilities worldwide. Ensuring the inclusion of persons with disabilities in the workforce is vital for the realization of their labour rights, including addressing discrimination and creating equal opportunities in the workplace (ILO, 2021).

The Crawling Framework (CF) we present here is an experimental implementation approach, a suite of tools dedicated to searching and applying for a job for people with disabilities. In this implementation we propose a crawler with modern architecture based on intelligent agents, a multi-agent system trained on the basis of a rule engine. We also present a case of using CF in a CV creation expert. It can be seen that our approach is perfectible by adding learning to the agent model.

II. Experimental work

Conceptually the CF follows the general characteristics and approaches of a multi-agent system with WebCrawler Softbots working on the Internet, exploring web pages to collect info on job offering, following links and pattern matching. The design approach used is a goal-based agent which downloads and indexes content from all over the Internet and automates the process of finding job listings across numerous websites and job boards.

For the main dashboard the access is restricted, only authorized users are allowed to use the application (Fig.1).

The primary actions in a job search include (Fig.2):

- ✓ *Define the job criteria:* To conduct a successful search, the first step is to select the job criteria, including

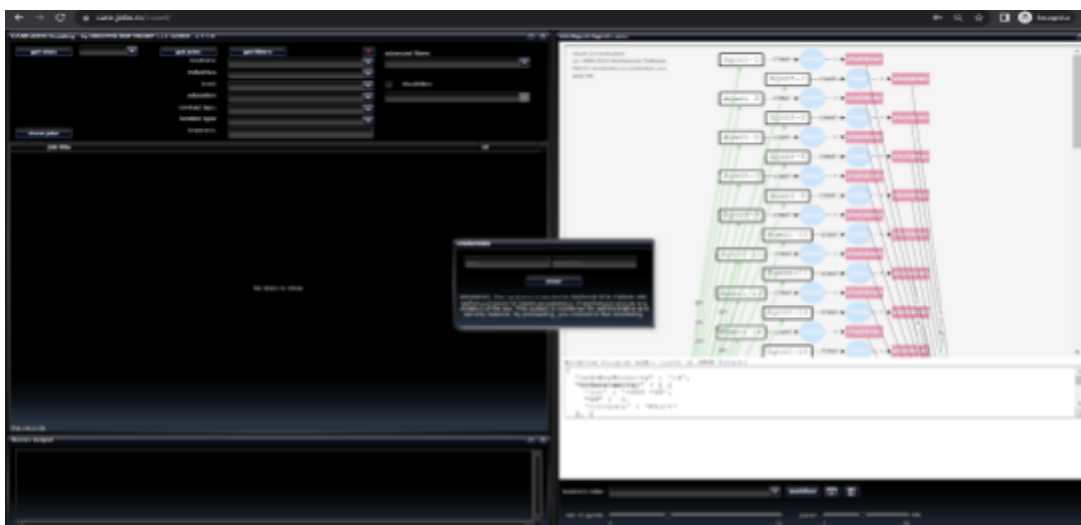


Figure SEQ Figure * ARABIC 1. CAREJOBS Crawling Framework – main dashboard

job type, location, industry, required qualifications and other details, such as employing people with disabilities.

- ✓ *Scanning job boards:* The Softbot can be configured to browse firm career pages and well-known job boards relevant to the industry in order to look for job openings.
- ✓ *Extracting job offers:* The Softbot extracts the relevant information from job listings, such as job title, company, location, job description and application instructions. It also captures the URL for the full job posting.
- ✓ *Filtering and sorting:* The results can be arranged and sorted according to our preferences with the help of intelligent agents. For example: relevance, date, location or accessibility for people with disabilities. The ultimate goal of the framework is to provide the CV Expert user with smart filters – or what we like to call advanced filters – based on complex criteria.

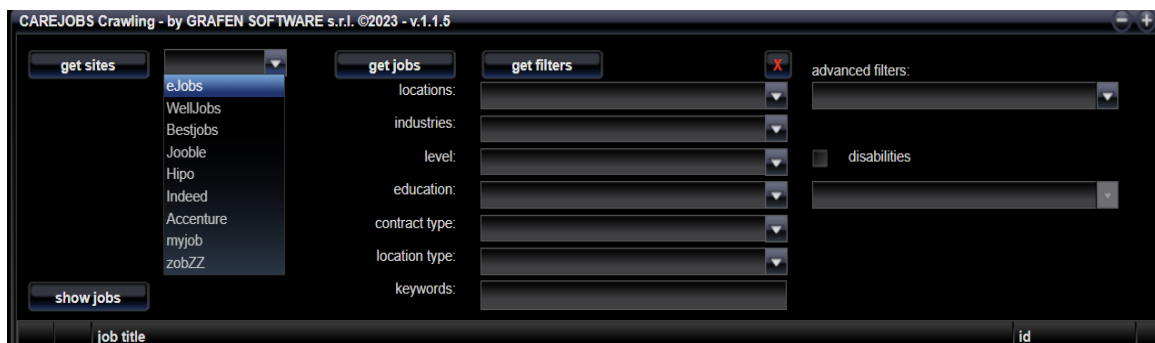


Figure 2. CAREJOBS – job searching

Our Web Crawler bots can exhibit intelligent agent-like behaviour (according to FIPA, 2022):

- ✓ *Task Automation:* Web Crawler bots automate tasks of repetitive nature, such as information retrieval, collection of large amounts of data, and monitoring. They can exhibit human-like behaviour in a way that they follow pre-programmed rules and instructions to execute tasks autonomously. Our distributed agents are scheduled overnight with a precise set of tasks through a Rules Engine.
- ✓ *Adaptive Behaviour:* Our intelligent agents are permanently aware of the environment they function in and the tasks at hand. They adapt to changes in the structure of websites they scrap while crawling. They can dynamically change their scraping patterns and rules.
- ✓ *Decision-Making:* Our Web Crawler bots are programmed to choose which links to follow or what data to extract based on specified rules.
- ✓ *Information Processing:* Our Web Crawler bots perform necessary information processing, such as data aggregation and categorization. They can collect data from many sources and structure it in a useful way to the user.
- ✓ *Interaction with Users:* Our intelligent agents interact with users to capture their preferences and requirements by filtering keywords in the CAREJOBS CV Expert application.

However, it's important to note that our WebCrawler softbots are typically rule-based and deterministic, which means their "intelligence" is limited to the rules and algorithms provided by the programmers. So there is no Learning, limited Adaptive Behaviour and limited Decision-Making due to the time and resources constraints of the project. An example of a desirable goal would be integration into chatbots or CV Expert UI, allowing users to provide feedback and receive online information. (Avram, D. 2023).

Our WebCrawler softbots for job searches respect the terms of service and privacy policies of the websites you're scraping. Some websites prohibit web scraping, so it's important to be aware of their policies and use the softbot ethically and responsibly. We use the robots.txt file that instructs the agents on what are the crawling permissions. Some sites are protected from DDOS attacks and may interpret crawling as that. We detect that and take care of the pace of crawling, speed, length, hits/seconds and other aspects of crawling.

Framework Components

Although the functionality of the agents is completely automated, almost all career offering search steps have a manual counterpart in the user interface to allow our experts to view results, configure, fine tune the processes and understand the dynamic of these processes. We call that CF Monitoring Application (CF-MA) which includes a series of dashboards with a portal layout for various tasks providing a deep visibility into the results and performance of the framework (Fig.6). The high-level architecture of the framework is represented in the diagram

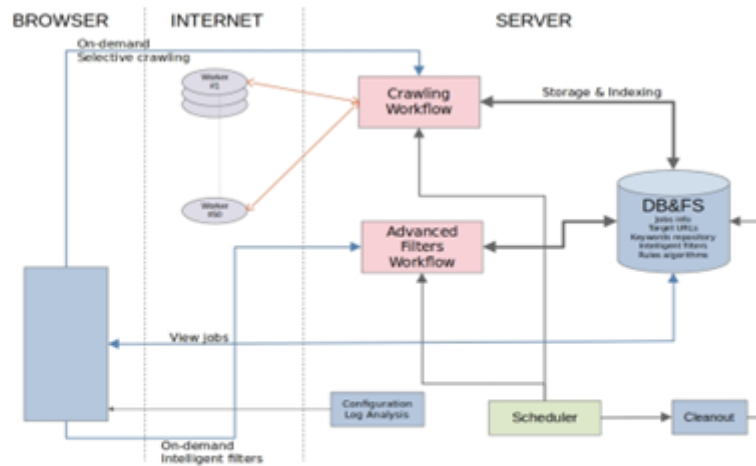


Figure 3. The CAREJOBS Crawling Framework Architecture

(Fig.3). The backbone of the crawling agents is a traditional Scientific Workflow with worker nodes to execute workflow tasks, following loosely the WS-BPEL 2.0 standard. The worker nodes are managed by a Rules Engine which is loosely based on Rule Interchange Format (RIF) standard under development at W3C.

We use 2 workflows:

1. for orchestrate WebCrawler softbots
2. for building intelligent filters from either job database or through crawling.

Worker nodes are orchestrated by a Rules Engine. The rules are written using YAML, a human-readable data serialization language. CF is a mission-critical application and the process of maintaining business logic within the source code can become too complicated so the Business Rules are used to ease the maintenance by separating business logic from the source code, giving the researchers and the system administrators the ability to handle fairly complex rules. The application provides an editor and a compiling tool to build new rules or modify the existing ones (Fig.4).

The CF-MA has also a dynamic representation (Fig.5) of the workflow showing live how WebCrawler bot work. This is beyond the spectacular show of how internal pieces are moving but it is an instrument to help configuring and fine tuning the workflows.

```

rule:
  id: 4
  name: "Disabilitati locomotorii"
  description: "Disabilitati locomotorii"
  status: 0
  priority: 1
  condition: description.contains("programator", "locomotorii");
  actions:
    - job.insert(4);
  
```

Figure 4. The YAML rules editor

The CF-MA has also a dashboard for viewing (Fig.6), understanding and fine tuning the framework functionality:

- Views and maintains target URLs;
- View and analyse the job searches filters:
 - traditional filters: location, level, education, contract type, location type
 - varied keywords

- o intelligent filters – result of rule-based agents work.
- o filtering on different types of disabilities: vision impairment, deaf or hard of hearing, mental health conditions, intellectual disability, autism disorder, physical disability
- View jobs content applying various filters.

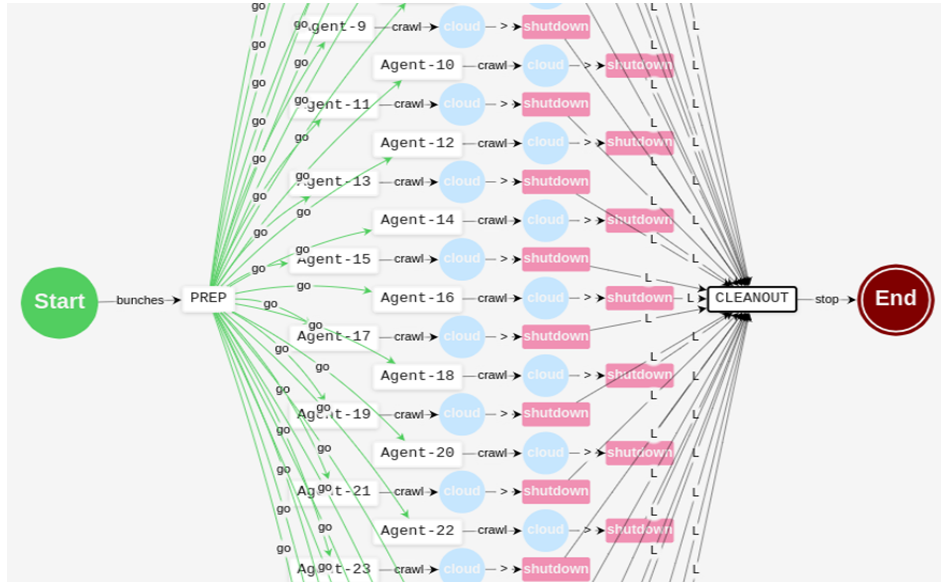


Figure 5. The CF-MA dynamic representation

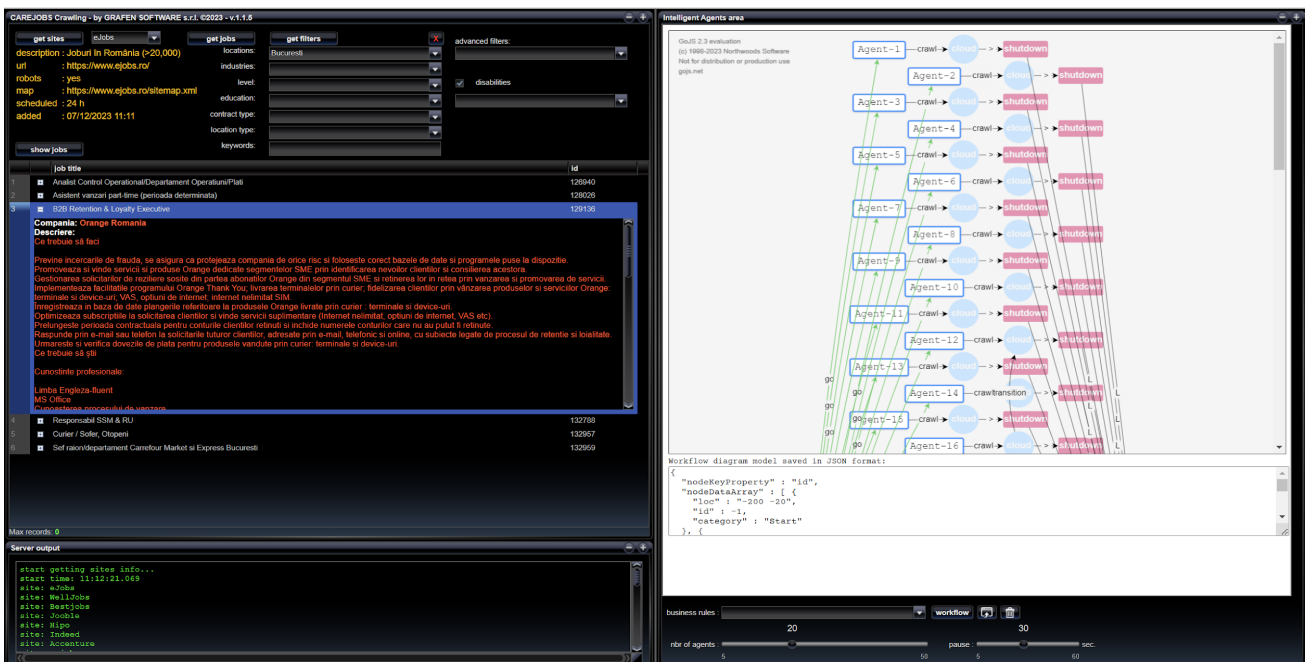


Figure 6. The main functionality of CF

Most of the functionality in this dashboard is exported through APIs into the CV Expert application in the CAREJOBS portal.

CF-API provides a RESTful service to access data about career offers to external users as well. It has the following characteristics:

- The service is SSL secured.

- The API requires authentication through an API key.
- The API has two components:
 - Get filters – to obtain the regular filters necessary to refine the search results – location, experience, studies, etc., and the advanced (intelligent) filters available at the moment.
 - Get jobs – to obtain career offering data for the required filters.

The API accepts *curl* calls for testing and getting to know the service.

III. Use case of CF

CV Expert (CVE) is an application launched from the CAREJOBS portal dedicated to helping disabled workers to build CVs using many interactive features and world-wide available resources with the Job Search feature realized with the APIs provided by the Crawling Framework (CF).

Disability is an umbrella term that covers any mental or physical impairment, from chronic diseases to physical injuries to mental disorders. Since disabled workers are not legally obliged to mention their disability on the CV, in a similar way the employers may hide jobs availability for disabled persons, so the ability to do a comprehensive search on the Internet for the disabled suitable jobs is the key to explore the career offering world. This is realized with the APIs provided by the Crawling Framework (CF) for Pre-Employment Processes which is a part of the CAREJOBS project. The CF framework provides:

- ✓ Access to Jobs Database.
- ✓ View jobs content applying various filters:
 - traditional filters: location, level, education, contract type, location type
 - input varied keywords
 - intelligent filters – result of the rule-based agents' work.
 - filtering on different types of disabilities: vision impairment, deaf or hard of hearing, mental health conditions, intellectual disability, autism disorder, physical disability etc.

Access the actual job hosting sites to view the jobs and eventually apply to them.

Figure 7 shows the search jobs window layout and how results are displayed to the end user.



Figure 7. CF integration on CVE

IV. CONCLUSION

Experimental development has proven the usefulness and efficiency of the crawling tool for finding specific jobs for people with disabilities. Intelligent agents and software architectures for intelligent agents were used to develop the framework. The research perfected the author's previous research, and not only that, validated by practical applications. Our approach is perfectible by adding learning to the agent model.

Also, following the results obtained through the experimental development of CAREJOBS crawling, the development team decided to use formal methods to verify and optimize the behavior of the agents, the formalization being able to help highlight unwanted behavior, as well as to find errors of omission.

Funding Statement

The research is carried out within the project "CAREJOBS - Resource platform for people with disabilities" SMIS 156142.

Conflict of interest

The author declares that there is no conflict of interest.

References

1. Avram, D. (2023). "Intelligent agents used in the digitization of pre-employment processes for people with disabilities", VII. Congress of International Applied Social Sciences - C-iasoS 2023. Valletta -Malta, 13-15 November 2023.
2. FIPA. (2022). "Document Set Standard FIPA Specifications, IEEE Foundation for Intelligent Physical Agents (FIPA)", IEEE Computer Society, <http://www.fipa.org/repository/standardspecs.html>.
3. ILO, (2021). A joint publication by Fundación ONCE and the ILO Global Business and Disability Network, developed within the framework of Disability Hub Europe, a project led by Fundación ONCE and co-funded by the European Social Fund, "An inclusive digital economy for people with disabilities". https://www.ilo.org/global/topics/disability-and-work/WCMS_769852/lang-en/index.htm
4. Lin, H. (2007). "Architectural Design of Multi-Agent Systems: Technologies and Techniques", IGI Global. <https://doi.org/10.4018/978-1-59904-108-7>
5. Zofia Wrona, Wojciech Buchwald, Maria Ganzha, Marcin Paprzycki, Florin Leon, Noman Noor, Constantin-Valentin Pal (2023), "Overview of Software Agent Platforms Available in 2023", <https://doi.org/10.3390/info14060348>