

THE ULTIMATE GUIDE TO

CV/Resume Parsing



daXtra

Introduction

Here at Daxtra, we get plenty of **questions about CV/resume parsing**. Most of these questions focus around some common themes, including what CV/resume parsing actually is, why it's important and what good parsing software looks like.

These are all valid questions. To help you answer some of these, we've put together a comprehensive guide on the subject. It covers these **common questions** and gives you an in-depth resource to refer back to whenever you need it.

In this guide, we'll cover:

1. What is CV/resume parsing?

2. The different types of CV/resume parser

3. The benefits and the challenges of CV/resume parsing

4. How to choose a CV/resume parser

But before jumping into the topic of CV/resume parsing, it's important to have a clear understanding of some of the common terms and abbreviations that are used when discussing CV/resume parsing. The next page lists out the terms to know.

Terms to know

Applicant Tracking System (ATS) — a software application that manages the entire process of recruiting a candidate, including managing job adverts and applications.

Candidate Relationship Management (CRM) — CRM stands for Candidate Relationship Management. It's a technology recruitment professionals use to manage interactions with candidates and potential candidates. CRMs are used to store past, current and potential candidate data and help recruiters and hiring managers deliver personalised experiences to candidates across all stages of the hiring process.

Resume — a summary of a candidate's education, qualifications, skills and previous experience, and other relevant information.

CV (Curriculum Vitae) — a summary of an individual's education, qualifications, skills and previous experience, and other relevant information. Sometimes used interchangeably with "resume", a CV is typically longer in length and is used in the UK, Australia and New Zealand.

Job Specification — a statement of the qualifications, personal characteristics, a brief description of the job and the skills required by an individual to perform the job in question.

Extensible Markup Language (XML) — a text-based format or metalanguage which allows users to define their own customised markup languages, especially in order to display documents on the internet. XML is one of the formats, alongside JSON, that unstructured CV/resumes and job descriptions are converted into during the parsing process.

JavaScript Object Notation (JSON) — is a text-based format that's used for representing structured data based on JavaScript object syntax. It is commonly used for transmitting data in web applications (e.g., sending data from the server to the client, so it can be displayed on a web page, or vice versa). JSON is one of the formats, alongside XML, that unstructured CV/resumes and job descriptions are converted into during the parsing process.

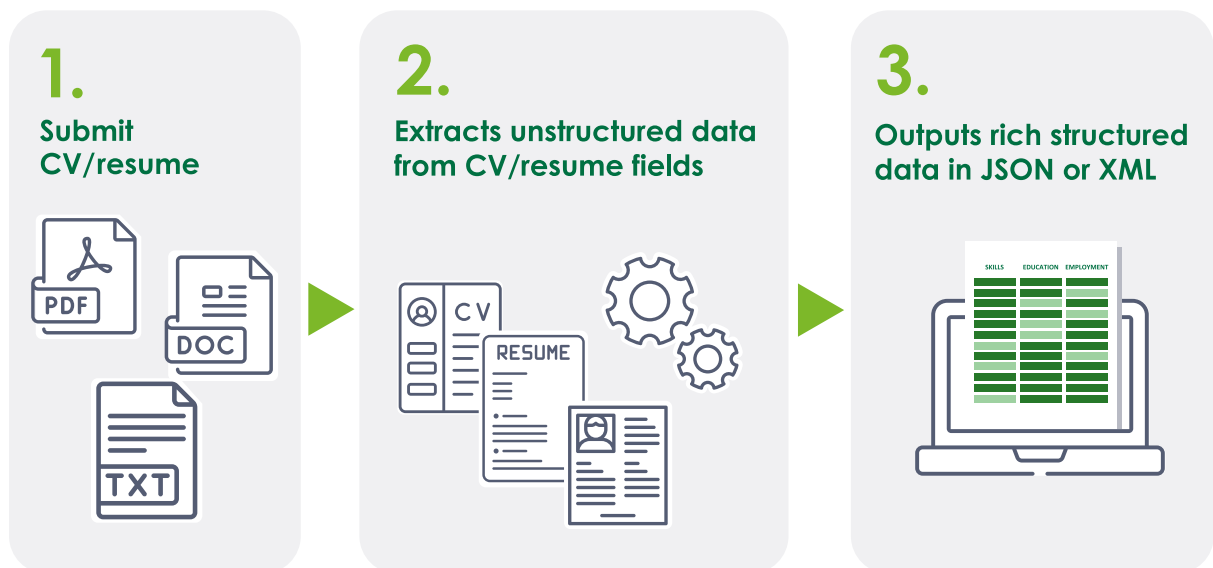
1. What is CV/Resume parsing?

CV/resume parsing, also known as CV/resume extraction, is the process of software **analysing** and converting unstructured CV/resumes and job descriptions from formats such as PDF, Microsoft Word Documents, Excel and Raw Text files into structured XML or JSON data. This conversion ensures your incoming documents are ready to load into another application such as an ATS or CRM.

What does a CV/Resume parser do?

CV/resume parsing software typically extracts information such as a candidate's professional skills, work experience, education history, contact details and achievements from their CV/resume. Parsing software can also be used to extract data from job descriptions such as the job title, job type, location of the position, salary details, contact information, required qualifications and skills.

A high-performance parser will detect and provide data from more fields than a low-performance one. For example, Daxtra's parser detects and extracts data from over 150 document fields and converts it into structured XML or JSON data. This parsing process helps to ensure that important candidate information is converted into a format that's ready for loading and storage within a database.



2. Types of CV/Resume parser

In general, there are **four main types of parser**. These are:

#1 Keyword-Based Parsers

A keyword-based parser works by identifying words, phrases and simple patterns in the text of a CV/resume. This is the simplest but least accurate CV/resume parser with only about a 70% accuracy rate. Keyword-based parsers are the least accurate because a keyword parser can't extract information that is not surrounding one of its keywords.

#2 Grammar-Based Parsers

A grammar-based parser uses grammatical rules to understand the context of every word in the CV/resume. Grammar-based parsers tend to capture much more detail than keyword-based parsers. Through computational semantics, grammar-based parsers can distinguish between different meanings when one word or phrase might have different contexts.

Grammar-based parsers can achieve accuracy rates well above 90%, but the downside of grammar-based parsers is that they require a lot of manual encoding by skilled language engineers. A substantial amount of testing is also required to make sure that improvements in one area do not degrade performance in another.

#3 Statistical-Based Parsers

A statistical-based parser applies grammar with a probability or finds the most probable parse of a sentence. A statistical-based parser can distinguish between contexts of the same word or phrase and can also capture a wide variety of structures such as addresses and timelines. For highest accuracy, it requires an input of a vast number of CV/resumes that are manually marked up with the information to be extracted.

#4 Hybrid Parsers

Some parsers, like Daxtra's, are a combination of two **different parsers**.

Daxtra parser is a hybrid of a grammar and statistical-based parser. This gives Daxtra's technology the best of both worlds – the powerful high accuracy of the grammar-based parser combined with the continual machine learning capabilities of the statistical parser. This means Daxtra's parsing technology is not only extremely accurate but also improves over time.

3. The benefits and challenges of CV/Resume parsing

Benefits

Extracts accurate data

Over the years, as parsing has developed, parsing accuracy has reached “near-human accuracy”. Parsing software such as Daxtra's is estimated to have achieved an accuracy level of around 90%. If a parser's accuracy is less than this, the number of errors it produces will be too large for it to function correctly without extensive human supervision.

Saves time

Parsing software allows recruitment teams to import incoming CV/resumes into their ATS in minutes. Rather than a recruiter manually extracting information, parsers extract candidate data from CVs/resumes and jobs, and convert the data into structured formats such as XML or JSON so that it's ready to be loaded and stored in a CRM. This eliminates the need for manual input, saving recruitment teams a considerable amount of time. Our customers find that Daxtra's parser automatically extracts CV/resume data that would typically take a human 10-15 minutes to extract manually.

Supports recruiter productivity

With tedious and manual tasks taken off their plate, recruiters can work smarter, are more likely to be engaged in their work, complete tasks more efficiently, and spend more time on higher-value tasks such as building relationships with candidates and clients.

Helps ensure a fairer hiring process

Recruitment teams use CV/resume parsing to streamline the applicant screening process. It's **estimated that around 98% of Fortune 500 companies** use CV/resume parsing within their ATS to support their hiring processes. Parsing also helps reduce ethical issues in hiring, such as unconscious bias.

A 2014 study conducted by research firm Isync Surveys and recruitment company Hays sent CV/resumes to 1,029 hiring managers, with the name being the only difference. Half the hiring managers received a CV/resume for Simon Cook, while the other received a CV/resume for Susan Campbell.

The survey results found that Simon was **more likely to get a callback**. In this study, the name on the CV/resumes was the factor causing the unconscious bias in the hiring process.

Parsing software can overcome problems like this by being programmed to extract all the candidate information from a CV/resume but only display objective candidate information such as a candidate's relevant skills and work experience to a hiring manager and use this information as the basis for fairer hiring decisions.

Challenges

Understanding context

If a parser doesn't fully understand the **context** of the terms used in a CV/resume, this could lead to the wrong information being parsed and important information being missed or miscategorised. The terms used to describe skills and qualifications listed on a CV/resume are constantly evolving, so high-quality parsing software should continually learn new terms and skills as the relevant CV/resumes are parsed.

This ability to learn new terms and skills and understand the context that they are being used in is typically seen in parsers that utilise machine learning such as Daxtra's.

Cost

The pricing for parsing software is typically determined by the number of documents you'll need to parse or the number of users that'll be using the software. Parsing software will often be more cost effective and produce more accurate results than an individual or a team parsing CV/resume and job data manually.

Keyword stuffing

Some CV/resume parsers can be open to manipulation by particular candidates through keyword stuffing on a CV/resume. If a candidate "stuffs" their CV/resume with the right keywords, they could make themselves appear a better fit for a posted job. However, a parser such as Daxtra's is not affected by keyword stuffing as it recognises skills in a CV/resume based on context.

4. How to choose a CV/Resume parser

There are a few factors **to take into consideration** when you're looking for a **high-quality parser**, such as:

Does the parser integrate with your CRM/ATS or other recruitment automation tools?

Most CRM and ATS solutions come with built-in parsing software. If the ATS or CRM you're using doesn't contain parsing software, you'll need to look at purchasing and testing parsing software as a component that can integrate with the CRM/ATS you're using.

How much time can parsing software save you in comparison to manual processes?

The parsing software you implement should save you a considerable amount of time compared to extracting information from a CV/resume manually. For example, high-performance parsers can extract the most complex of CV/resumes within 1-3 seconds,

How accurate is the parser?

As we mentioned earlier, if a parser is less than 90% accurate, the number of errors will be too large to permit it to load data into a CRM or ATS without extensive human supervision. Aim to make sure the parsing software you're using is at least 90% accurate. Once again, you can do this by testing the accuracy of the parsing software by running a number of documents through it and monitoring the accuracy of the data that is parsed from both pieces of software and checking to see whether the data extracted has been mapped to the relevant places inside your CRM.

If you're a CRM provider and looking to test parsing software before integrating it into your platform make sure you test the accuracy of the parser and the speed at which it can extract data. Typically, a parser that extracts a substantial amount of information from a CV/resume will be slower than a parser that extracts less data due to the amount of data that is extracted.

Can it parse in multiple languages?

It's common for most parsers to parse in a few different languages, and it's best practice to have an idea of the different languages you'll need CV/resume and job data to be parsed in before purchasing parsing software. At Daxtra, our parsing software parses in **over 40 languages**, and our team

of language engineers deliver regular engine updates, which continually improve native accuracy across all languages. To find the best candidates, wherever they may be in the world – and to compete internationally, **multilingual CV/resume software** is essential.

Does the parsing software support various formats?

Candidate CV/resumes and job descriptions come in various formats, so a state of the art parser should be able to support all popular input formats, including Doc, Docx, HTML, RTF, PDF, PNG and JPEG and convert these into output formats such as structured XML or JSON.

Does the parser understand human context?

It's important that the parser you use understands the context of the information on a **candidate's CV/resume**. For example, parsing software needs to understand if 'MD' on a CV/resume means 'Medical Doctor, Maryland or Managing Director'. A good CV/resume parser will use machine learning, **natural language processing** and semantics to detect the relevance of what's expressed on a CV/resume.

Conclusion

This guide should have given you a good idea of what CV/resume parsing is, the benefits a high-quality parsing solution can bring to recruitment teams and how to go about choosing a parser for your business. If you're looking into the possibility of implementing parsing software into your tech stack, remember to take the points we've covered in this guide into consideration.

Take the time to:

- **Test and measure the accuracy of a parsing solution**
- **Measure the length of time it takes for the parsing software to extract information from a CV/resume**
- **Measure how much time an effective parsing solution can save your recruitment or talent acquisition teams.**

You'll then be equipped with the key facts you need to choose and implement an effective CV/resume parser into your recruitment tech stack.

At Daxtra, **our parsing software** is consistently benchmarked as the **most accurate parsing software** on the market and helps over 2500 clients worldwide extract and convert candidate and job information in 40+ languages from over 150 data fields. If you have further questions about parsing, how our parser works or how it can help your business, do **get in touch**.