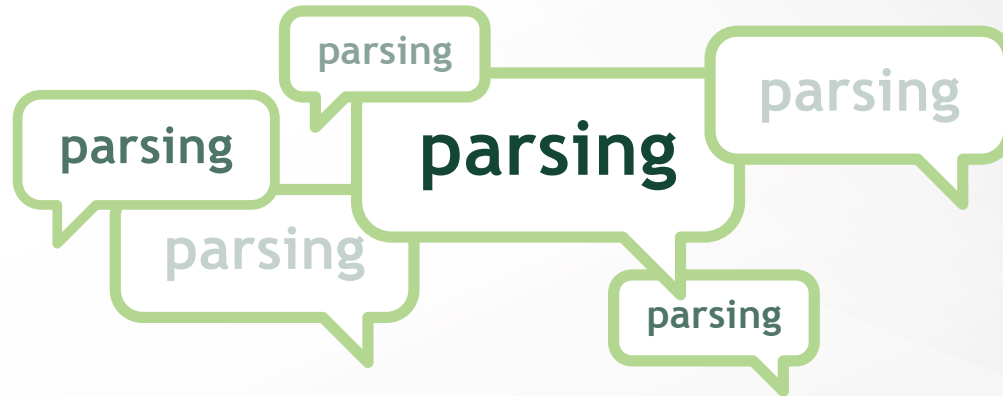# Types of Resume Parsers
and How They Work

Types of Resume Parsers and How They Work

As online automation progresses 'parsing' is a word you're going to hear more frequently.

In recruiting and staffing we've all seen and heard about resume or CV parsers or job parsers and how they automate the recruitment workflows with quick, accurate results. But not all parsers are the same. Let's explore the different types of resume parsers and what they are capable of.

Contents

What is a Parser?

First let's start by defining parsing. Essentially, it means extracting meaning from a sequence of words.

Parsing usually applies to text and the act of reading natural language, analysing, and converting the text to computer language. This is where we get the term Natural Language Processing (NLP), which is part of the AI functionality in parsers.

According to G2 Learning Hub, a resume parser automatically extracts, stores and analyses resume or candidate data in a way that the information is then able to be categorised, sorted and most importantly, searched.

Recruitment software industry leader Bullhorn describes the three main benefits of resume parsing as being:

- ✓ increased speed
- ✓ improved candidate experience
- ✓ better data in your ATS

The Three Approaches

In general, there are three approaches to parsing a resume or job for use by Applicant Tracking System (ATS) or Candidate Relationship Management (CRM) software.

Keyword-Based

Grammar-Based

Statistical-Based

Keyword-Based

A keyword-based parser works by identifying words, phrases and simple patterns in the text of a resume.

Then a simple _heuristic algorithm_, designed to solve a problem in a faster and more efficient fashion than traditional methods, is applied to the text around these words.

This algorithm may look for something that looks like a postal code and then try to interpret the surrounding words as an address. Or it may search for patterns that look like date ranges and assume the surrounding text is an employment timeline.

This is the simplest but least accurate kind of resume parser with only about a

70%

accuracy rate.

It is less accurate because a keyword-based parser can't extract information that is not surrounding one of its keywords. If a keyword is too vague such as the skill 'director' then the parser will frequently make the wrong guess about its interpretation because the word or phrase is too broad a term to accurately identify.

Grammar-Based

A grammar-based parser uses grammatical rules to understand the context of every word in the resume.

It also combines words and phrases together to make complex structures that capture the meaning of every phrase in the resume. These grammar rules or language restrictions form a hierarchy of formal languages which is known as the Chomsky Hierarchy, after 'the father of modern linguistics' Noam Chomsky. In the late 1950s Chomsky explored the rules of grammar in his linguistic theories on sentence structure and syntactic constraints.

Much more complex than a keyword-based parser, a grammar-based parser, or grammar as they are called, generally captures much more detail. Through computational semantics, grammars are also capable of distinguishing between different meanings when one word or phrase might have different contexts.

With grammar-based parsers it is possible to achieve accuracy rates well above

VS

Near human accuracy which is rarely greater than

90%

96%

The downside is a grammar-based parser requires 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.

Statistical-Based

A statistical-based parser applies grammar rules with a probability or finds the most probable parse of a sentence.

Like a grammar-based parser, a statistical-based parser can distinguish between different contexts of the same word or phrase and can also capture a wide variety of structures such as addresses and timelines. It applies numerical models of text to identify structure in a resume.

To be most accurate, it requires an input of a vast number of resumes that are manually marked up with the information to be extracted. The more resumes they have statistics on, the better the data.

While it performs better than a keyword-based parser, a statistical parser is not as accurate as a grammar-based parser on data the parser has not been trained on. For a statistical-based parser to be accurate, it has to be previously trained on the data it is expected to process. This is where machine learning comes into play.

As defined in SAS Insights,

"Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention."

Inevitably, the more data available, the more accurate the outcome.

Hybrids

Some parsers can be a combination of these approaches,
or rather a hybrid of more than one kind.

DaXtra's parser for example, is a combination of a grammar and a statistical parser.
It possesses the best of both parsers — the powerful, high accuracy of the grammar-
based parser combined with the continual machine learning functions of the statistical
parser. It is not only extremely accurate, but also improves over time. With the
millions of resumes DaXtra parses monthly, the accuracy level continues to grow.
It is ranked as the industry-leading parser with near-human level accuracy.

Parsers as a Tool

A resume and job parser is an important part of the process that helps increase recruitment productivity and eliminate human bias.

This technology allows recruiters to locate, gather, store and organise data in applications, resumes or jobs.

Once acquired, this data is easily searchable using keywords or Boolean search terms. Advanced search and match technology is able to search and rank candidates by job experience, relevant skills and educational background. Parsing technology has advanced so much using NLP and machine learning that some resume parsing programs achieve what is called 'near human accuracy' with up to 95% efficiency.

Instead of manual data entry and sorting through hundreds of resumes daily, the speed and accuracy of parsing technology allows recruiters to focus on the more important tasks like one-on-one meetings and phone calls. It can also create a better candidate experience with quicker response times.
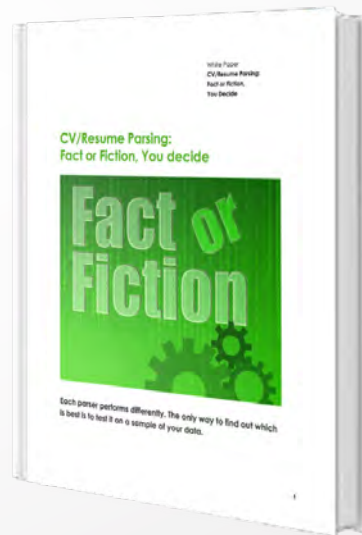
Ultimately, resume parsing automation is a great tool that recruiters can use to make their jobs easier. It can also increase a company's ROI by not only bringing in more qualified candidates, but in leveraging the candidate information in its own database.

Additional white papers you may find of interest
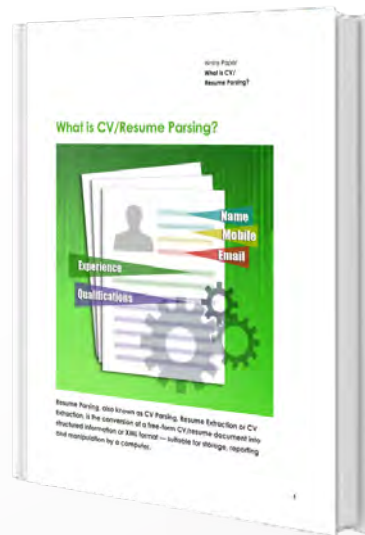
What is
CV/Resume Parsing?

CV Parsing Fact or Fiction:
You Decide

Key Measurements
of a Good Parser

What is
Semantic Search

About Us

DaXtra Technologies is a world-leading specialist in high accuracy multilingual resume and job parsing, as well as semantic search, matching and aggregation technologies.

We automate the entire employment application process, helping you quickly source the most relevant candidates.

Using DaXtra's software is a great way to save on the administrative overheads of retrieving, optimising, storing, managing, analysing and searching candidate and job data.

Stay ahead of the competition by finding candidates faster with DaXtra's intelligent automated recruitment solutions.

EMEA:
London, UK: +44 20 7801 6323
info_uk@daxtra.com

US:
East Coast (Richmond, VA): 804 767-1351
West Coast (Berkeley, CA): 925 262-1174
info_us@daxtra.com

APAC:
Hong Kong: +852 3695 5133
Australia: +61 2 8705 5445
Japan: +81 3 6868 8734
China: +86 21 8025 1557
apac@daxtra.com

www.daxtra.com