Types of Parsers and How They Work

Posted February 26th, 2014

In general, there are three types of approach to parsing a CV/resume:

- **Keyword based parsers**

- **Grammar based parsers**

- **Statistical parsers**

Keyword-based parsers

Definition.- A keyword-based CV parser works by identifying words, phrases and simple patterns in the text of the CV/Resume and then applying simple heuristic algorithms to the text they find around these words. This are the simplest and least accurate kind of CV parser.

Features.- These tools may look for something that looks like a postal code and then try to interpret the surrounding words as an address, or they may look for patterns that look like date

ranges and assume that the surrounding text is an employment timeline.

Accuracy rate.- It is **hard to get beyond 70% accuracy**. This type of CV parser is the least accurate because they can't extract information that is not surrounding one of their keywords, and if their keywords are ambiguous (e.g. the skill "Director") then they will frequently make the wrong guess about its interpretation.

Grammar-based parsers

Definition.- Grammar-based ones contain an enormous number of grammatical rules that seek to understand the context of every word in the CV/resume. These same grammars also combine words and phrases together to make complex structures that capture the meaning of every sentence in the resume.

Features.- These parsers are much more complicated than keyword-based parsers, and generally capture much more detail and are also capable of distinguishing between the different meanings that one word or phrase might have within different contexts.

Accuracy rate.- It is possible to achieve **accuracy rates well above 90%** (human accuracy is rarely greater than 96%). The downside is that this type of resume parser requires a lot of manual encoding by skilled language engineers, and a lot of testing to make sure that improvements in one area do not degrade performance in another.

Statistical parsers

Definition.- This type of parser attempts to apply numerical models of text to identify structure in a CV/Resume. Like grammar-based parsers, they can distinguish between different contexts of the same word or phrase and can also capture a wide variety of structures such as addresses, timelines, and the like.

Features.- To be most accurate, they require as input a vast number of CV/Resumes that are manually marked up with all the information that is required to be extracted.

Accuracy rate.- This kind of parser usually performs better than a keyword-based one, but not so well as grammar-based parsers on data that the parser has not been trained on. Thus, for a statistical parser to be accurate, it has to be previously trained on the data that it is expected to process.

So, what are the key measurements of a good CV parser?

Tags: DaXtra Blog, parsing cv, parsing resume, resume parser, cv parser, cv parsing

f 🐦 in

Aashma Uprety
27/09/2018 à 05:49:15

How to parse CV using python?I need help for my final year project.Please kindly reply.

Reply to *Aashma Uprety*

cwatson
01/10/2018 à 15:13:05

Please send us an email with your question, to: info_uk@daxtra.com. This way it will get directed to the proper people to answer your question! Thanks for reaching out and good luck!

Reply to *cwatson*

First Name *

Last Name

Email *

Website

Comment *

protection par **reCAPTCHA**

Confidentialité - Conditions

Submit Comment

Blog Categories

- DaXtra Blog
- News
- Events
- Resources
- Case Studies
- White Papers

Learn more about our intelligent recruitment solutions.

CONTACT US TODAY!

Home
DaXtra Capture
Daxtra Search
DaXtra Parser
DaXtra Magnet
DaXtra Components
Join Daxtra
Daxtra Blog

Accreditation

Connect with Us

SUPPORT   **CAREERS**   **CONTACT US**

Recruitment Management Software   Resume Management Software   Automated Recruiting Software
CV Matching Software   Resume Parsing Software   Applicant Management System
CV Management Software   Applicant Management Software   Recruiting Management Software

© 2020 Daxtra Technologies Ltd

Privacy Policy