



Data Assessment Solutions GmbH  
Misburger Str. 81b  
D-30625 Hannover  
+49 511 47 4023 30  
info@data-assessment.com  
[www.data-assessment.com](http://www.data-assessment.com)

## Profil-Parser

---

Die Data Assessment Solutions GmbH (DAS) entwickelt eine Softwareplattform, auf der Unternehmen die Fähigkeiten (Skills), Verfügbarkeiten und Projekteinsätze ihrer Mitarbeiter managen können. In der Anwendung werden unter anderem Mitarbeiterprofile gepflegt, in denen Skills, Zertifizierungen, Projekterfahrungen und weitere Informationen dargestellt werden. Diese Profile sind Lebensläufe, die mit zusätzlichen Details angereichert werden. Neue Mitarbeiter müssen diese Daten initial im System erfassen. Ein Teil dieser Informationen ist in der Regel bereits in anderen Quellen, z.B. im LinkedIn-Profil des Mitarbeiters, vorhanden. Um diese Inhalte einfach übernehmen zu können, müssten sie zunächst in eine geeignete Struktur überführt werden.

Ziel dieses Projektes ist die Entwicklung eines Software-Tools, mit dem sich Lebenslaufdaten aus einem Dokument extrahieren und in ein einfach verarbeitbares Format konvertieren lassen.

Im ersten Schritt sollen die gesuchten Informationen aus einem LinkedIn-Profil extrahiert werden. Aus Sicherheitsgründen soll das Tool nicht selbst auf LinkedIn zugreifen. Stattdessen wird die Profilseite im Web-Browser durch den Nutzer aufgerufen und per Copy & Paste übertragen. Um den Prozess einfach zu gestalten, wird die gesamte Seite markiert (Strg-A in Windows), kopiert und in ein Eingabefeld des Software-Tools eingefügt.

Das Verfahren über Copy & Paste hat den Vorteil, dass es sich einfach auf andere Quellen, z.B. andere Webseiten oder Word-Lebensläufe, erweitern lässt.

Das Frontend des Tools soll als Web-Oberfläche umgesetzt werden. Um eine visuelle Kontrolle der übertragenen Seite zu ermöglichen, soll diese in einen Rich Text HTML Editor kopiert werden. Hierfür kann ein frei verfügbarer Editor, wie z.B. CKEditor, verwendet werden.

Aus der übertragenen HTML-Seite müssen die relevanten Profildaten identifiziert und extrahiert werden. Hierzu zählen:

- Name
- Stadt / Region / Land
- Profilfoto
- Firma und Position
- Zusammenfassung
- Berufserfahrungen (vorherige Firmen)
- Ausbildung
- Skills

Alle HTML-Formatierungen außer Aufzählungen und Zeilenumbrüche werden bei der Extraktion entfernt. Die erhaltenen Daten sollen in einer geeigneten Struktur gespeichert werden.

Die Ergebnisse sollen dem Benutzer in der Oberfläche des Tools übersichtlich angezeigt werden. Das Tool transformiert damit den vorgegebenen HTML-Inhalt in eine Standardstruktur, die nur die gewünschten Informationen enthält.

Als Ausbaustufe können Profildaten aus anderen Quellen in das Tool kopiert und geparkt werden. Besonders interessant ist die Verarbeitung von Word-Lebensläufen, wie sie für Bewerbungen verwendet werden. Im Gegensatz zu LinkedIn-Profilen hat man hier keine einheitliche Struktur für alle Profile. Lebensläufe sind individuell gestaltet, was die Identifikation der oben aufgelisteten Inhalte erschwert.

Die Implementierung des Tools sollte in Python oder C# erfolgen. Nach Absprache können auch andere Programmiersprachen verwendet werden.

Wir freuen uns auf euch und eine tolle Zusammenarbeit!

Johanna de Vries

[devries@data-assessment.com](mailto:devries@data-assessment.com)

## Über Data Assessment Solutions

*Seit ihrer Gründung 2008 beschreitet Data Assessment Solutions neue Wege für die effiziente Nutzung von Skills und Mitarbeiterkapazitäten in dynamischen Wachstumsbranchen – insbesondere für IT-Services- und Consulting-Unternehmen. Wir vereinen langjährige Branchenerfahrung mit tiefer Expertise in maschinellem Lernen und Datenanalyse. Durch die intelligente Nutzung vorhandener Daten erreichen wir deutliche Effizienzsteigerungen und liefern unseren Kunden eine verlässliche Quelle für alle Informationen zu Mitarbeitern und Projekten.*