**Data exploration**

***A data exploration of LinkedIn profile data***

*8vance Matching Technologies BV*

*Venlo*

|  |
| --- |
| **Date : 25-02-2016** |
| **Version : 0.1** |
| **Status : Concept** |
| **Document name : Data exploration v0\_1** |
| **Author : Tim Hermens** |

**Version**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version** | **Date** | **Author(s)** | **Changes** | **Status** |
| 0.1 | 25-02-2016 | Tim Hermens | First concept. | Concept |
|  |  |  |  |  |

**Table of contents**

1. Introduction 4

1.1 Goal of this document 4

1.2 Punctuation 4

# Introduction

## Goal of this document

The company 8vance Matching Technologies BV uses scraping techniques on various social media websites liked LinkedIn to collect a large amount of profile data of companies and people. The problem is that this data often misses interesting (parts of) information. The data especially lacks a complete list of skills of employees/unemployed (this is called the KSC data).

The goal of this document is to analyze the profile data of employees/unemployed. The result of this analysis will be documented within this document. The goal of the analysis is to find interesting data and/or patterns that can be used to create links between the data. E.g. a large percentage of the people who followed education A have skills A, B and C. So a link has been found from education A to skills A, B and C (and not vice versa).

## Punctuation

**,**

bla bla

**&**

bla bla

**.**

bla bla

**/**

bla bla

**-**

bla bla

**:**

The colon is needs to be parsed as follows:

|  |  |  |
| --- | --- | --- |
| **Context** | **Priority** | **Action** |
| [i]:[1] | 1 | If a string is like ii:2 or 2:ii, the colon needs to stay. So if there's an i on the left side of the colon, only i's can be on the left side. If there's a number, only numbers can be there. |
|  |  |  |
|  |  |  |
|  |  |  |

**@**

bla bla

**;**

bla bla

**+**

b

**=**

b

**|**

b

**!**

b

**\***

b

**`**

b

**#**

b

**%**

b

**<**

b

**?**

b

**\_**

b

**\**

b

**>**

b

**~**

b

**$**

b

**^**

b