**Automatically match people with jobs**

Enrich personal data provided by people to create better matches



|  |
| --- |
| **Author : Tim Hermens** |
| **Place : Blerick** |
| **Date : 02-05-2016** |
| **Version : 0** |

**Student information**

Name: Hermens T.  
Student number: 2210369  
Graduation course: Software Engineering (full-time)  
Graduation period: 15-02-2016 to 24-06-2016

**Company information**

Name: 8vance Matching Technologies BV  
Department: Research & Development  
Location: Blerick, Kazernestraat 17

**Company mentor information**

Name: Keuren P.  
Function: Neural Software Developer

**School mentor information**

Name: Schouten G.

**Thesis information**

Title: Automatically match people with jobs  
Date of issue: -  
Confidential: Strictly confidential (no distribution allowed)

**Signed and approved by company mentor**

Date: -  
Signature:

|  |
| --- |
|  |

# Preface

This thesis contains a description of the project I've executed for the company 8vance Matching Technologies BV. This company is active in the data science area (also known as Big Data) of the IT industry. One of their products is called AIMA, which is a digital agent that is able to match profiles of people with jobs. The main objective of my project is to improve the quality of matches of people with jobs by enriching the profile data of people. This project ran from February until June 2016.

This project has been one of the most challenging projects I've done, mainly because of my limited experience with the data science area. With help and insights from the company's data science experts (Sabrina Ziebarth, Lou Cremers, Paul Keuren, Jan Jacobs) I managed to overcome many challenges, for which I would like to express my gratitude. I want to thank Paul Keuren in particular for his thorough feedback and excellent support. And last but by no means least, I would like to thank Gerard Schouten for his close involvement and continuous support throughout the whole project.

# Table of contents

[Summary 5](#_Toc449962913)

[Samenvatting 6](#_Toc449962914)

[Glossary 7](#_Toc449962915)

[1 Introduction 8](#_Toc449962916)

[2 The company 9](#_Toc449962917)

[3 The assignment 10](#_Toc449962918)

[4 The approach 11](#_Toc449962919)

[5 Orientation phase 12](#_Toc449962920)

[6 Research and solution phase 13](#_Toc449962921)

[7 Completion phase 14](#_Toc449962922)

[8 Conclusion and recommendations 15](#_Toc449962923)

[Evaluation 16](#_Toc449962924)

[Bibliography 17](#_Toc449962925)

[Appendices 18](#_Toc449962926)

[A. Project Initiation Document (PID) 18](#_Toc449962927)

[B. Other appendices 18](#_Toc449962928)

# Summary

*(An informative summary with a maximum length of ONE page, summarizing the whole thesis.*

*Describe the company and the problem that needed to be tackled.*

*Describe the chosen approach with arguments.*

*Describe the results.*

*Describe the conclusions and recommendations.*

*DON'Ts:*

*Don't refer to other chapters of the thesis.*

*Don't go too in-depth on details.*

*Don't use difficult terms somebody may not understand. Anyone should be able to read the summary.*

*Don't add images or bulleted lists.)*

# Samenvatting

*(The same rules apply from the previous chapter.)*

# Glossary

**AIMA (Automatic Intelligent Matching Agent)**

This is one of 8vance's products that's able to match people with jobs.

**ME (Matching Engine)**

This is one of 8vance's products that can find similarities between profiles and create matches between the most similar profiles. For instance, it's able to match jobs and talents for 8vance's AIMA product, and it's able to mach real estates and buyers for 8vance's Domizz product.

*(Define unknown words, terms, symbols, abbreviations in alphabetic order. Define abbreviations as follows:*

*<Full term> (<Abbreviation>): <Definition>*

*Explain all the things defined in the glossary in the thesis' text as well when they're first encountered.)*

# Introduction

*(Start with a nice, catchy phrase that stimulates further reading. Think of interesting questions, a catchy phrase, a comparison/formula, an incredible number, etc.*

*Describe global information of the project. Describe the company, the problem, the assignment and relevance of the assignment. Also briefly mention the approach/strategy.*

*The final paragraph will explain the document's structure and what information can be found in every chapter. You can also explain the structure of individual chapters.)*

# The company

## Foundation and mission

8vance Matching Technologies BV is a relatively young and innovative organisation in the data science (also known as Big Data) area of the IT industry. The company was founded at November 2012 with the mission to render recruitment and career development smart, social, simple, cost-effective and fast. There're three main reasons why the company wants to invest to accomplish this mission.

Firstly, unemployment still remains in the top 10 world problems (Hutt, 2016). The partial cause of this problem is the fact that vacancies are spread all over the internet on a variety of websites (e.g. online job boards or a company's website) and the unemployed have trouble finding appropriate vacancies. The company sees an opportunity here to collect as many vacancies and profiles of people as possible from a variety of websites and store them on one central place. Since finding a perfect job (or employee for a company's recruiter) is a challenge in itself, the company wants to provide assistance in the form of an Automatic Intelligent Matching Agent (AIMA) that's able to match people with jobs. This should help to decrease the unemployment rate.

Secondly, the company distinguishes itself from the competition by offering an even smarter, state-of-the-art matching engine. The company will be the first to combine both hard and soft skills to improve the matching quality between people and jobs.

And lastly, the matching engine that'll be developed to accomplish the mission can be used to accomplish a wide variety of incredible and innovative things (yet to be discovered). For instance, it would be possible to create a career assistance agent that's able to offer suggestions of skills you should achieve and/or educations you should follow to make progress in your career development.

## Products

The company is working on several products. The following products are the three main products they're working on:

* **AIMA**. An automatic intelligent matching agent that's able to match people with jobs. Since the project is focussed on this product, it'll be discussed more in-depth in the upcoming section.
* **Sjerlok**. An artificial intelligence system that uses Big Data to track down stolen goods for Delta Lloyd. (Vonk, 2015)
* **Domizz**. An online platform for buying and selling real estates. An automatic digital real estate agent finds matches between estates and interested buyers.

Because of the small number of employees and strict deadlines of the AIMA and Sjerlok products, the company had to cease the development of Domizz. However, an ex-CEO of www.jaap.nl is now working for 8vance and has taken up the responsibility to continue the development of Domizz.

### AIMA

As said before, AIMA is an automatic intelligent matching agent that's able to match people with jobs. This service is available on 8vance's website (www.8vance.com) after registering an account for free.

The target audience of this product are companies and talents[[1]](#footnote-1) around the world. The product will mainly be used by the companies' recruiters and the talents.

The general use of the product consists of the following three steps. (8vance Matching Technologies, 2016)

#### Step 1 - Company profile and vacancies

Companies can register with 8vance through a online wizard. The wizard gathers the required information about the company it'll need for its matching algorithm. A short leadership test is also part of the registration, which improves the results of the matching algorithm. The company can start posting vacancies of jobs when it's registered. The wizard assists step by step in the creation of a full job DNA. This way AIMA gathers all the required information to find better matches (talents).

#### Step 2 - Matching and scouting

AIMA uses several methods to find talents. When a company activates one of their vacancies, AIMA searches the internet for public CVs to find talents (this is called scouting). She will establish who has worked where for how long and which competences have played an important role. AIMA may automatically add missing competences which are in line with a talent's working experience and educational background to improve the quality of matching (this part is the goal of this project).

In addition to publicly available information, talents can also create their own 8vance profile. Recruiters have the possibility to upload their database of talents and have AIMA match these data as well. AIMA creates a list of matches containing information about the found job matches or talent matches.

#### Step 3 - Advertising and social media

In a few steps, companies can create creative and efficient online campaigns for their vacancies, which includes a link to their own home page. Companies can also create advertisements on job sites and social media. Talents who have seen the advertisements and are interested, can register with 8vance so that they're immediately matched to the vacancy in question.



Figure - 8vance's headquarters in Blerick in The Netherlands

## Organisation

The company currently consists of a total of 15 employees. An organisation chart of this company can be seen in <Figure> below.

<The organisation chart>

The company is situated in the Netherlands and Romania. The software development team is situated in Romania and the data science and marketing teams are situated in the Netherlands. I'm part of the data science team in the Netherlands.

### Manufactuur

As can be seen in Figure 1, the company's headquarter is situated in the Manufactuur building in Blerick in the Netherlands. The Manufactuur used to be one of the fabric halls of Leolux (a company that specializes in design furniture), but is now an open and transparent work floor for promising start-up companies. The Manufactuur currently houses over five companies, such as Accerion, Yubu and 8vance. (Kickstart Venlo, 2016)

*(A detailed description of the company. When, where and why has it been established? Where are they now? Are there any sister companies? What is their position in the market? How many employees (organisation chart?)? What are they building and for whom?*

*Give examples of products they create. Also explain in which section of the company I am working.)*

# The assignment

## Project goal

The company is working on a product called AIMA that matches jobs and talents as accurately as possible. Another product called ME (Matching Engine) is responsible for actually creating the matches between profiles. Essentially, AIMA uses the ME to create the matches between jobs and talents.

The ME uses a variety of algorithms to create the matches. The algorithms need to be trained with data initially for them to be able to create matches. The algorithms are trained on company, job and talent data that 1) are scraped from social network websites such as LinkedIn and Xing, and 2) are retrieved from 8vance's internal database. One major problem is that the talent data from the first data source misses useful information more than 98 percent of the time.

The information that's missing varies from talent to talent. The missing information can consist of followed educations, work experience, owned skills, the work industry, dates, and more. The most notable missing data in this list are the owned skills, because the matching quality is the most dependant on the skills a talent owns. The more skills are missing, the higher the likelihood of a mismatch between job and talent. This doesn't necessarily result in the failure of the ME, for it has found the best matches based on the incomplete information. But it does result in the failure of AIMA to find the best matches between talents and jobs, because the lack of skills leads to an incomplete picture of the talent's skill set which severely limits the possibilities of matches. Therefore the focus and the goal of this project lies on complementing the skills of the scraped talent data to increase the quality of matches.

The preferred solution to this problem comes in the form of an algorithm that's able to complement the missing skills for the provided talents. This solution can be used in a stand-alone tool to provide skill suggestions on request, and can be integrated in the scraping process to automatically complement the skills for newly scraped talent data.

## Scope

The scope of this project is specified in <table>, which is taken from the appendix Project Initiation Document (PID).

## Constraints

* Because the scraping of the talent data is done in a Python application and the solution must be able to be integrated in this application, means the solution must be compatible with Python.
* The stand-alone tool that's developed must be platform-independent.
* A software architecture document must be

*(Provide a detailed description of the assignment. Explain:*

* *The initial situation. What's the problem? Why is that the problem? What are the consequences?*
* *What's the project's goal? What are you and the company trying to achieve? What's the preferred solution?*
* *Define the precise assignment description.*

*Describe the scope and constraints of the project (mandatory languages, documents, methods like scrum/prince?).*

*Describe any changes during the project.*

*Describe the research questions and research framework.*

*Be concrete in the assignment/problem specification. Don't say "This caused delays.", but say "This caused forty percent of the customers to receive their orders four months later.")*

# The approach

(Start with a introduction, explaining the details of this chapter.

Describe the chosen methods I've used during the project (phasing, agile, tsp) and why. Also mention the research methods (inquiries, interviews, proof of concepts, architecture) and research strategies (field, library, lab, etc.) with explanation.

Describe relevant topics from the PID like the project planning and contact moments.)

# Orientation phase

(Start with a introduction, explaining the details of this chapter.

Topics of importance, in this order:

* PID. Why is it useful?
* Initial research. Mention its results and why this was done. This includes profile data analysis and scraping method analysis.

End with a conclusion, briefly mentioning the results of the topics and the next steps to take.)

# Research and solution phase

(Start with a introduction, explaining the details of this chapter.

Topics of importance, in this order:

* Requirements specification. Explain why and mention a range of important and unimportant requirements (for competence reasons).
* Software architecture document. Explain why (kruchten, ISO 25010) and refer to a simplified model of the chosen architecture (must be simple enough to understand for anyone). Also mention the architectures that were also considered. Also mention the things of the architecture I will be responsible for to implement and the things 8vance will be responsible for.
* Building the algorithm library. Started with it after the architecture was partially created. This library contains four major features: pre-processing the data, using an algorithm to predict skills, and post-processing the data.
  + Pre-processing of the data. Explain why it's crucial. This includes the degree model (interview), using existing taxonomies (interview), and normalizing the data. Also mention the problem of creating an own model and the lack of a 'major' model.
  + Algorithms to predict skills. Mention the algorithms that were considered and tested. Explain why one algorithm was better than the other and which one was the best. Explain how the algorithms were tested and briefly explain how the best algorithm works.
  + Post-processing of the data. Explain why it's crucial. Maybe I'll leave this out because there isn't much interesting to say about this because it's very similar to the pre-processing section.
* Building the algorithm analysis "tool". Mention the problem with the first architecture of this tool (a complete GUI application with a 3-layer architecture) and why this architecture wasn't necessary. Explain the use of this tool and why it's important.

End with a conclusion, briefly mentioning the results of the topics and the next steps to take. Important questions to answer are:

* Does the found solution solve the problem?
* Is the found solution satisfactory to the company?)

# Completion phase

(Completion phase might not be the best translation of "invoeringsfase", but it will do for now.

Start with a introduction, explaining the details of this chapter.

Topics of importance, in this order:

* Conclusions and recommendations of the research. Briefly mention the best found solution and its flaws. Mention recommendations for the company to get possibly better results.
* Finalization of the architecture document. (especially the deployment view which isn't created yet) What part of the architecture is and isn't implemented?
* List of implemented, partially implemented and not implemented requirements.

End with a conclusion, briefly mentioning the results of the topics. A conclusion may be unnecessary for this chapter though.)

# Conclusion and recommendations

(The reader must be capable to understand this chapter without reading any of the intermediate chapters. This chapter cannot suddenly introduce new information out of nowhere.

Determine whether or not the project has been successful by reviewing the found solutions. Mention recommendations to improve the solutions.)

# Evaluation

(Use of "I" is mandatory in this chapter.

Reflect on your own work process and experiences. Describe what I've learned, what I enjoyed, and what were the most important learning moments. Emphasize especially how I solved my mistakes.

Also mention and reflect on my personal learning goals.)

# Bibliografie

Hutt, R. (2016, January 21). *What are the 10 biggest global challenges?* Opgeroepen op May 8, 2016, van World Economic Forum: https://www.weforum.org/agenda/2016/01/what-are-the-10-biggest-global-challenges/

(APA notation literature reference list)

# Appendices

## Project Initiation Document (PID)

(Contains the whole PID document.)

## Other appendices

(This section is just a placeholder for other appendices. Only add documents or preferentially parts of documents if they offer relevant information for the reader.)

1. Talents are people who are looking for a job. [↑](#footnote-ref-1)