Introduction
Server's part
Matching Readme's to Job Ads
Client side

Radboud University Nijmegen

Selecting job advertisments towards a GitHub accounts

Rafal Klimek, Sven van Valburg, Agnieszka Rakowska

Radboud University Nijmegen

Software Analysis project presentation
24th June 2016

Introduction
Server's part
Matching Readme's to Job Ads
Client side

Radboud University Nijmegen

Table of Contents

Introduction
Server's part
Matching Readme's to Job Ads
Client side

Radboud University Nijmegen

How we do that

1. Analyze README files from Github accounts
2. Search advertisement from Indeed platform
3. Matching developers experience and jobs advertisemens
4. Counting results

Introduction
Server's part
Matching Readme's to Job Ads
Client side

Radboud University Nijmegen

Table of Contents

Introduction
Server's part
Matching Readme's to Job Ads
Client side

Radboud University Nijmegen

How we implemented server

1. NODE.js
2. Asynchronous requests to gather data

Introduction
Server's part
Matching Readme's to Job Ads
Client side

Radboud University Nijmegen

Table of Contents

Introduction
Server's part
Matching Readme's to Job Ads
Client side

Radboud University Nijmegen

Algorithm Overview

1. Use DBpedia to find all key terms from readme's and job ads
2. Assign a weight to these terms
3. Create vectors for the user and all job ads
4. Compare these vectors using the weights (Cosine Similarity)

Introduction
Server's part
Matching Readme's to Job Ads
Client side

Radboud University Nijmegen

TF-IDF

- Term Frequency - Inverse Document Frequency
- TF, the more a word appears in a text the higher the weight
- IDF, The more documents a word appears in, the lower the weight
- Eventual weight : $TF \cdot IDF$
- Our metric: $NormalizedTF \cdot (1 + log(N/n))$

Introduction
Server's part
Matching Readme's to Job Ads
Client side

Radboud University Nijmegen

Table of Contents

Introduction
Server's part
Matching Readme's to Job Ads
Client side

Radboud University Nijmegen

How we implemented API

1. Form with three fields
2. Table with results ordered with similarity with Github users profile and advertisements from Indeed