Combining Deep Learning Matching and Traditional Keyword Search for Job Search

Henning Rode

Luigi Lorato, Eike Dehling, Vincent Slot, Mihai Rotaru

textkernel
Machine Intelligence for People and Jobs
About Textkernel

Founded in 2001
Headquartered in Amsterdam
> 125 employees

International leader in AI for HR and Recruitment
More than 1000 clients worldwide
Part of the CareerBuilder group
Textkernel’s Match

The traditional approach
From a Job Text to a Query

Sr. Optical Network Engineer - Central and East Regions

WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

RESPONSIBILITIES:
* Perform physical site readiness assessments
* Work with customer engineering teams to secure change windows for optical network installation or migration
* Successfully install optical network equipment [..]

QUALIFICATIONS:
* Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years) [..]
* Experience with fiber optic characterization (FOC) a plus.
* Transport technologies (Ethernet, DWDM, ROADM)
WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

**RESPONSIBILITIES:**
* Perform physical site readiness assessments
* Work with customer engineering teams to secure change windows for optical network installation or migration
* Successfully install optical network equipment [...]

**QUALIFICATIONS:**
* Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years) [...]
* Experience with fiber optic characterization (FOC) a plus.
* Transport technologies (Ethernet, DWDM, ROADM)
Sr. Optical Network Engineer - Central and East Regions

WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

RESPONSIBILITIES:
* Perform physical site readiness assessments
* Work with customer engineering teams to secure change windows for optical network installation or migration
* Successfully install optical network equipment [...]

QUALIFICATIONS:
* Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years) [...]
* Experience with fiber optic characterization (FOC) a plus.
* Transport technologies (Ethernet, DWDM, ROADM)
Sr. Optical Network Engineer - Central and East Regions

WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

RESPONSIBILITIES:
* Perform physical site readiness assessments
* Work with customer engineering teams to secure change windows for optical network installation or migration
* Successfully install optical network equipment

QUALIFICATIONS:
* Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years)
* Experience with fiber optic characterization (FOC) a plus.
* Transport technologies (Ethernet, DWDM, ROADM)

Textkernel's parser extracts from the text:
- Job title
- Education
Sr. Optical Network Engineer - Central and East Regions

WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

RESPONSIBILITIES:
* Perform physical site readiness assessments
* Work with customer engineering teams to secure change windows for optical network installation or migration
* Successfully install optical network equipment […]

QUALIFICATIONS:
* Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years) […]
* Experience with fiber optic characterization (FOC) a plus.
* Transport technologies (Ethernet, DWDM, ROADM)
Sr. Optical Network Engineer - Central and East Regions

WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

RESPONSIBILITIES:
* Perform physical site readiness assessments
* Work with customer engineering teams to secure change windows for optical network installation or migration
* Successfully install optical network equipment [...]

QUALIFICATIONS:
* Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years) [...]
* Experience with fiber optic characterization (FOC) a plus.
* Transport technologies (Ethernet, DWDM, ROADM)
Sr. Optical Network Engineer - Central and East Regions

WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

RESPONSIBILITIES:
* Perform physical site readiness assessments
* Work with customer engineering teams to secure change windows for optical network installation or migration
* Successfully install optical network equipment [..]

QUALIFICATIONS:
* Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years) [..]
* Experience with fiber optic characterization (FOC) a plus.
* Transport technologies (Ethernet, DWDM, ROADM)

Textkernel’s parser extracts from the text:
- Job title
- Education
- Experience years
- Skills
- Etc...

Enriched with Search Knowledge (synonyms, taxonomies, etc.)

The extracted query is used to search among the parsed CVs.
Sr. Optical Network Engineer - Central and East Regions

WTW has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

RESPONSIBILITIES:
* Perform physical site readiness assessments
* Work with customer engineering teams to secure change windows for optical network installation or migration
* Successfully install optical network equipment [...] 

QUALIFICATIONS:
* Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years) [...] 
* Experience with fiber optic characterization (FOC) a plus.
* Transport technologies (Ethernet, DWDM, ROADM)
Textkernel Match! Architecture Outline

CV parsing

job=Java Developer
langskill=German
experience=7
city=Utrecht

job=23
branch=IT
langskill=DE
experience=7
loc=(52.09-5.10)

Semantic understanding

Data Model of extracted and derived information.

Match!

Search!

Data Model

job branch
langskill
location
experience

Vacancy parsing

job=Java Ontwikkelaar
langskill=Duits
experience=7
city=Amsterdam

job=23
branch=IT
langskill=DE
experience=5.10
loc=(52.37-4.90)+30

Semantic understanding
Match query (AI)

Job title handling

• ‘Project Manager, Automotive’

Requiredness

• ‘Knowledge of Java is essential’

Balancing

• Learning a Weighting model
Important keyword extraction

Match in greater detail

- Java developer for financial services
- Open-world keyword extraction
  - Using gazetteers
  - And textual features
Perfect Match?
Perfect Match?
Not yet
Looking closer...

Sr. Optical Network Engineer - Central and East Regions

WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

RESPONSIBILITIES:
* Perform physical site readiness assessments
* Work with customer engineering teams to secure change windows for optical network installation or migration
* Successfully install optical network equipment [...]

QUALIFICATIONS:
* Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years) [...]
* Experience with fiber optic characterization (FOC) a plus.
* Transport technologies (Ethernet, DWDM, ROADM)
Looking closer... at the results...

Sr. Optical Network Engineer - Central and East Regions

WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

RESPONSIBILITIES:
* Perform physical site readiness assessments
* Work with customer engineering teams to secure change windows for optical network installation or migration
* Successfully install optical network equipment [..]

QUALIFICATIONS:
* Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years) [..]
* Experience with fiber optic characterization (FOC) a plus.
* Transport technologies (Ethernet, DWDM, ROADM)

SUMMARY

Over 4 years of extensive experience in the field of Networking with excellent written and communication skills. Strong understanding of Cisco routers, firewalls and switches and in-depth understanding of networking concepts, TCP/IP and communication protocols in LAN / WAN environments.

WORK EXPERIENCE
2013 – Present Network Engineer

* Designed, deployed and maintain the entire company’s network comprising of 6 branch offices, up to 500 employees.
* Configured and managed Cisco routers, Firewalls, layer 2 switches, Wireless and Web-content filter
* Performed periodic upgrades on all network devices in a timely manner.
* Configured Voice VLANs and QoS for VoIP.
* DHCP, DNS, SMTP and Kiwi syslog servers on Windows 2008 r2 server.
Looking closer... at the results... and the job

**Sr. Optical Network Engineer - Central and East Regions**

WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

**RESPONSIBILITIES:**
* Perform physical site readiness assessments
* Work with customer engineering teams to secure change windows for optical network installation or migration
* Successfully install optical network equipment

**QUALIFICATIONS:**
* Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years)
* Experience with fiber optic characterization (FOC) a plus.
* Transport technologies (Ethernet, DWDM, ROADM)

**SUMMARY**

Over 4 years of extensive experience in the field of Networking with excellent written and communication skills. Strong understanding of Cisco routers, firewalls and switches and in-depth understanding of networking concepts, TCP/IP and communication protocols in LAN / WAN environments.

**WORK EXPERIENCE**

2013 – Present **Network Engineer**

- Designed, deployed and maintain the entire company's network comprising of 6 branch offices, up to 500 employees.
- Configured and managed Cisco routers, Firewalls, layer 2 switches, Wireless and Web-content filter
- Performed periodic upgrades on all network devices in a timely manner.
  - Configured Voice VLANs and QoS for VoIP.
  - DHCP, DNS, SMTP and Kiwi syslog servers on Windows 2008 r2 server.
Limitations of the current approach

- **Complex pipeline:** impact of error propagation
  - job title extraction
  - taxonomy classification
  - synonyms
  - ...
Limitations of the current approach

- **Complex pipeline**: impact of error propagation
  - job title extraction
  - taxonomy classification
  - synonyms
  - ...
- **Fixed-term matching**: Stripping the text of semantic outside of extracted terms

**Sr. Optical Network Engineer - Central and East Regions**

WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

**RESPONSIBILITIES:**
- Perform physical site readiness assessments
- Work with customer engineering teams to secure change windows for optical network installation or migration
- Successfully install optical network equipment [...]

**QUALIFICATIONS:**
- Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years) [...]
- Experience with fiber optic characterization (FOC) a plus.
- Transport technologies (Ethernet, DWDM, ROADM)
Limitations of the current approach

- **Complex pipeline:** impact of error propagation
  - job title extraction
  - taxonomy classification
  - synonyms
  - …

- **Fixed-term matching:** Stripping the text of semantic outside of extracted terms
  - Semantic of extracted terms also limited to synonyms and taxonomy

**Sr. Optical Network Engineer**

- Central and East Regions

WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

**RESPONSIBILITIES:**

- Perform physical site readiness assessments
- Work with customer engineering teams to secure change windows for optical network installation or migration
- Successfully install optical network equipment …

**QUALIFICATIONS:**

- **Education:** Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years)
- **Experience:** Fiber optic characterization (FOC) a plus. Transport technologies (Ethernet, DWDM, ROADM)
Limitations of the current approach

- **Complex pipeline**: impact of error propagation
  - job title extraction
  - taxonomy classification
  - synonyms
  - ...
- **Fixed-term matching**: Stripping the text of semantic outside of extracted terms
  - Semantic of extracted terms also limited to synonyms and taxonomy
- **Skills**: Important, unimportant?
- **Other keywords**: too many unimportant
Limitations of the current approach

- **Overwhelming UX**
Deep Learning Matcher

The Neural IR approach
Can we build document embeddings from entire CVs and Job texts such that, within this continuous semantic space, looking for good matching jobs for a CV becomes a nearest neighbor search?
Could we use Doc2Vec? ... Not really


- Relevant CVs are close to each other
- But far from the job
Could we use Doc2Vec? ... Not really

- Relevant CVs are close to each other
- But far from the job

What’s the problem?
- Embeddings represent general language model
- Job and CV texts use different language:
  - E.g. ‘You should be familiar with ...’ vs ‘I have experience with ...’
  - Company information vs Personal Information
  - ....
- Differences get picked up in Doc2Vec
Starting from Word Embeddings can we capture only the relevant information for matching Jobs to CVs?
Extracting what’s relevant

Idea: Learn representation in shared space of CV and Job texts such that Relevant CVs are close to the Job and Irrelevant CVs are far.

- Convolutional Neural Networks extract from the Word Embeddings of the text the relevant parts
Extracting what’s relevant

Idea: Learn representation in shared space of CV and Job texts such that **Relevant CVs are close** to the Job and **Irrelevant CVs are far**.

- **Convolutional Neural Networks** extract from the Word Embeddings of the text the relevant parts
- **Contrastive Loss** models the space for relevant and irrelevant pairs as we need
Extracting what’s relevant

Idea: Learn representation in shared space of CV and Job texts such that Relevant CVs are close to the Job and Irrelevant CVs are far.

- Convolutional Neural Networks extract from the Word Embeddings of the text the relevant parts
- Contrastive Loss models the space for relevant and irrelevant pairs as we need
- Sharing the weights forces to extract only the shared relevant information for matching
That sounds great!
But where do we get the data?
Data

We train the model using Application data

- 10 Millions CV-Job application pairs
  - CV-Job pair **relevant** if it was a real application
  - Randomly sample **irrelevant** CV-Job pairs
Training

Using the application data we can now model the Job/CV vector space as we need and get job-meaningful neighbourhoods

Relevant pair cosine similarity
Irrelevant pair cosine similarity

Cosine similarity of Relevant and Irrelevant pair during training

before training
Looking at our Optical Network Engineer again...

Sr. Optical Network Engineer - Central and East Regions

WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

RESPONSIBILITIES:
* Perform physical site readiness assessments
* Work with customer engineering teams to secure change windows for optical network installation or migration
* Successfully install optical network equipment [...]

QUALIFICATIONS:
* Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years) [...]
* Experience with fiber optic characterization (FOC) a plus.
* Transport technologies (Ethernet, DWDM, ROADM)
...and at the most relevant candidate for DLM...

Sr. Optical Network Engineer - Central and East Regions

WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

RESPONSIBILITIES:
* Perform physical site readiness assessments
* Work with customer engineering teams to secure change windows for optical network installation or migration
* Successfully install optical network equipment […]

QUALIFICATIONS:
* Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years) […]
* Experience with fiber optic characterization (FOC) a plus.
* Transport technologies (Ethernet, DWDM, ROADM)

SUMMARY
* A dynamic professional in Telecommunication Engineering (SDH) in the field of O & M for Fiber Optics Communication Sector.

EXPERIENCE
Jan '04-2010 - Till Date with Unimax Telecom Pvt. Ltd. as a Network Engineer (O & M) Company Name: Unimax Telecom Pvt. Ltd,

* Knowledge of NMS (Tejas Networks), fault identification and rectification.
* Coordinating with the team in installing indoor and outdoor SDH, Ethernet & TDM equipment.
* External alarm monitoring and escalation of electrical works.
* Responsible for operation & maintenance of MAN network of TTSLS both for equipments and fiber. […]
SUMMARY
* A dynamic professional in Telecommunication Engineering (SDH) in the field of O & M for Fiber Optics Communication Sector.

EXPERIENCE
Jan '04-2010 - Till Date with Unimax Telecom Pvt. Ltd. as a Network Engineer (O & M) Company Name: Unimax Telecom Pvt. Ltd,

* Knowledge of NMS (Tejas Networks), fault identification and rectification.
* Coordinating with the team in installing indoor and outdoor SDH, Ethernet & TDM equipment.
* External alarm monitoring and escalation of electrical works.
* Responsible for operation & maintenance of MAN network of TTSL both for equipments and fiber. […]

Sr. Optical Network Engineer - Central and East Regions

WWT has an optical network engineering position available on the Professional Services team. As a Sr. optical network engineer, you will be responsible for the entire deployment lifecycle of an optical installations including design, installation, validation, and support.

RESPONSIBILITIES:
* Perform physical site readiness assessments
* Work with customer engineering teams to secure change windows for optical network installation or migration
* Successfully install optical network equipment […]

QUALIFICATIONS:
* Education: Bachelor's degree in a relevant engineering or technology field preferred. Strong technical background. (Minimum 10 years) […]
* Experience with fiber optic characterization (FOC) a plus.
* Transport technologies (Ethernet, DWDM, ROADM)
Evaluation

We compared the **DLM** with the **traditional Match**

- Look at 100 jobs
- and at the Top 5 CVs returned

What we got...
DLM is 33% less likely to return irrelevant results
UX Limitations of the Approach
The importance of being Explainable

- **Explainability**: User needs to have a clear understanding of the reason why they're seeing certain results.

- **Refinability**: User needs to be able to modify the generated query as preferred.

- **Control**: Having a faceted search system for filtering or searching on meta-data is crucial.
Combining DL Matching and Keyword Search
The importance of being Explainable

Hybrid approach: Traditional Match & DLM combined
In an ideal world

Query:

• Jobtitle = ‘Software Engineer’
• Skills = ‘Java’, ‘Elasticsearch’, ‘Python’
• Experience > 5 years
• Vector = (0.0, 0.3, 0.6, 0.0, 0.4, ...) [512 dimensions]
The challenge

- Combine search approaches
  - Fulltext & facets
  - Vector similarity

- Challenge
  - Search engine => text (‘software engineer’)
  - Vector similarity => numbers (0.0, 0.3, 0.6, 0.0, 0.4, …)
  - Reranking vs. Score combination
Approach 1: Keep it separate / Rerank
Approach 2: All-in-one

Indexing Jobs

1. Calculate embedding vector
2. Deep Learning Matcher
3. Store document + vector

Searching relevant Jobs

1. Calculate embedding vector
2. Search on fulltext and vector

Indexing Jobs/Resumes

Deep Learning Matcher

Fulltext Search (Elasticsearch)

Searching
## Pro’s / Con’s

<table>
<thead>
<tr>
<th>Separated</th>
<th>All-in-one</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pro’s</strong></td>
<td><strong>Pro’s</strong></td>
</tr>
<tr>
<td>● Off-the-shelf software</td>
<td>● Simpler searching / indexing</td>
</tr>
<tr>
<td>● Less engineering</td>
<td>● More robust</td>
</tr>
<tr>
<td></td>
<td>● Allows for score combination rather than only reranking approaches</td>
</tr>
<tr>
<td><strong>Con’s</strong></td>
<td><strong>Con’s</strong></td>
</tr>
<tr>
<td>● More parts, more complexity</td>
<td>● Need to engineer efficient vector-similarity in a search engine</td>
</tr>
<tr>
<td>● Complex two-phase search with difficult data transfer (execution speed)</td>
<td></td>
</tr>
<tr>
<td>● Need to keep both systems synchronized, updates!</td>
<td></td>
</tr>
</tbody>
</table>
Why Reranking is bad

1. Identify candidates by Keyword Search, rerank by DLM
   - Losing control:
     - only possible to add filters,
     - no way to add keywords, skills, job titles that would influence the ranking

2. Identify candidates by DLM, rerank by Keyword Search
   - Efficiency: Cannot prune search space by facet filters
   - Losing the fine-grained DLM relatedness score
Score Combination

Problem: Typical IR scoring functions (e.g. BM25) have no defined score range

• Textkernel scoring plugin guarantees [0 - 1] scores
• Score combination then is matter of weighting
Efficient nearest neighbour search in ES

Open Source Development to the rescue...

• Developed own ES plugin for vector-similarity search with reduced dimensionality -> losing precision
• Since ES 7.3: https://www.elastic.co/blog/text-similarity-search-with-vectors-in-elasticsearch allowing high-dimensional vector search
• It’s still limited in applicability and gets hopefully soon further extended
Meanwhile...

DL network is used to improve conventional architecture

• Improve quality of synonym knowledge
  • Check closeness of synonyms in vector space to filter out inappropriate ones
  • Build average CV embedding model for each skill, suggest skills with close embedding model.

• Job title normalization
  • Pre-filter: Make sure the normalized job title is also close in vector space. Mainly reducing embarrassing errors.
The End

Any questions?
Further Deep Learning Match Research

Applications are not symmetrical:

- Same job
- Promotion
- Side step

Tag application data accordingly before learning.

- Ignore promotion and side steps for learning the network to create embeddings.
- Promotion or Side-step models could be useful still for better matching.
Multilingual Fingerprints

Using existing knowledge to map English model to other languages, creating synthetic learning data.