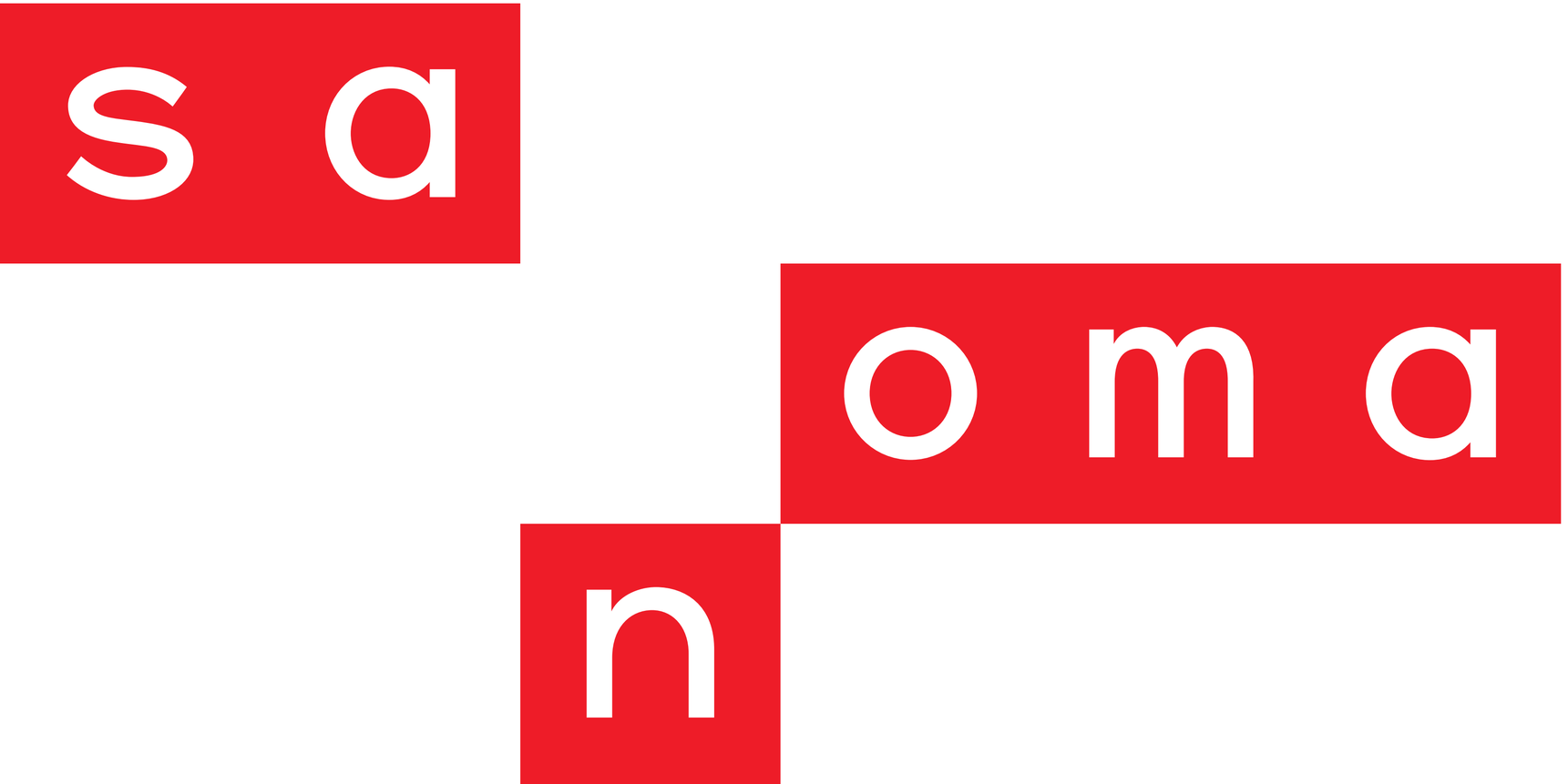
Matching Needs and Desires

Job and Candidate Matching using Machine Learning



Clemens Westrup
@ICMA 12.10.2017

Hello!

Who I am

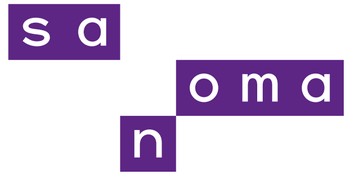
Clemens Westrup (firstname.lastname@sanoma.com)

Data Scientist at Sanoma, a media and learning company

Background: Computer Science, Machine Learning and Product Development

This talk

Matching Jobs and Candidates using Machine Learning



Overview

20 min

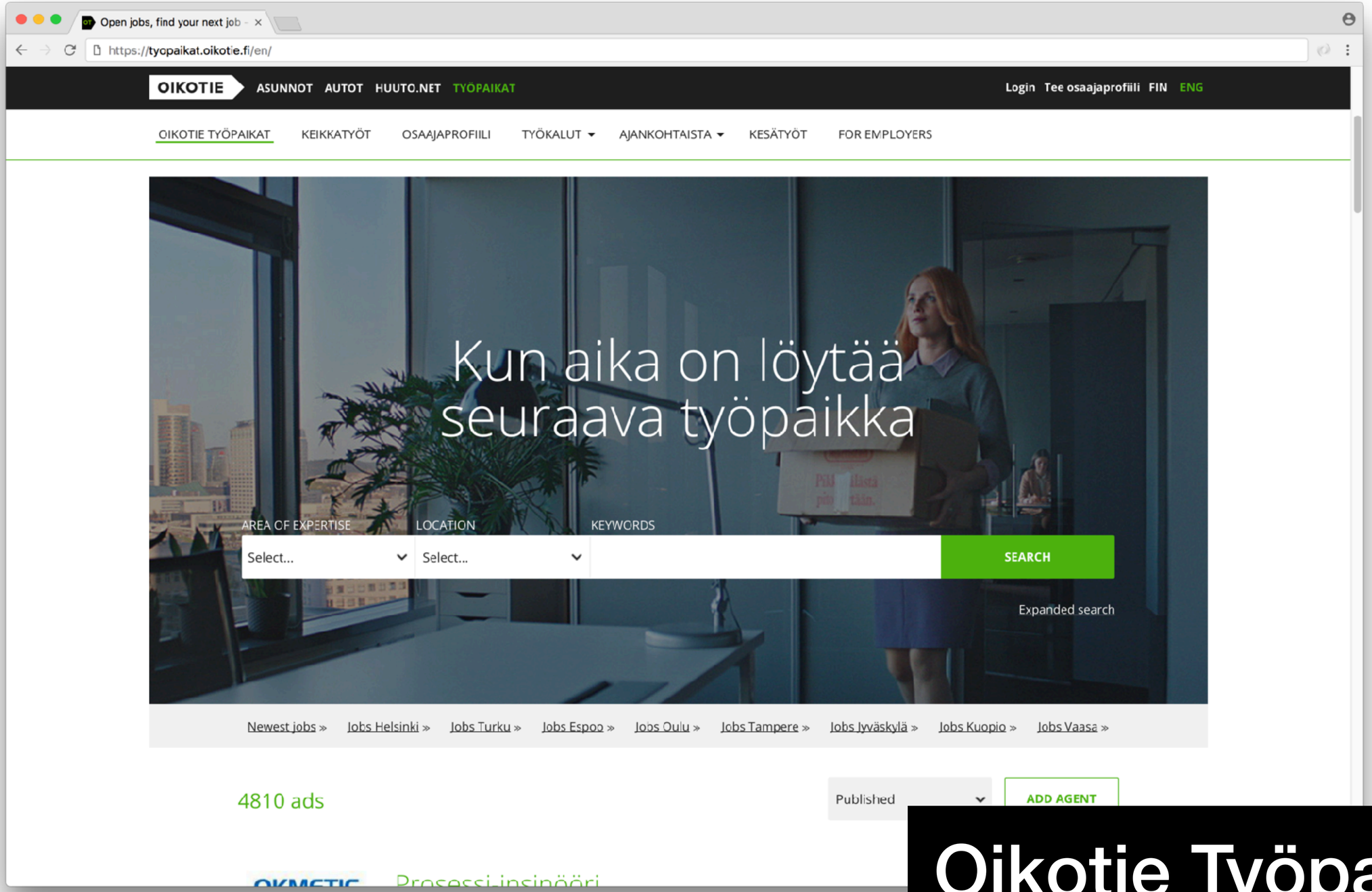
1. Intro: Job Recommendations at Sanoma
2. The Challenge: Matching Jobs and Candidates
3. Approaches: A Small Deep Dive into Machine Learning
4. Learnings and Takeaways

10 min

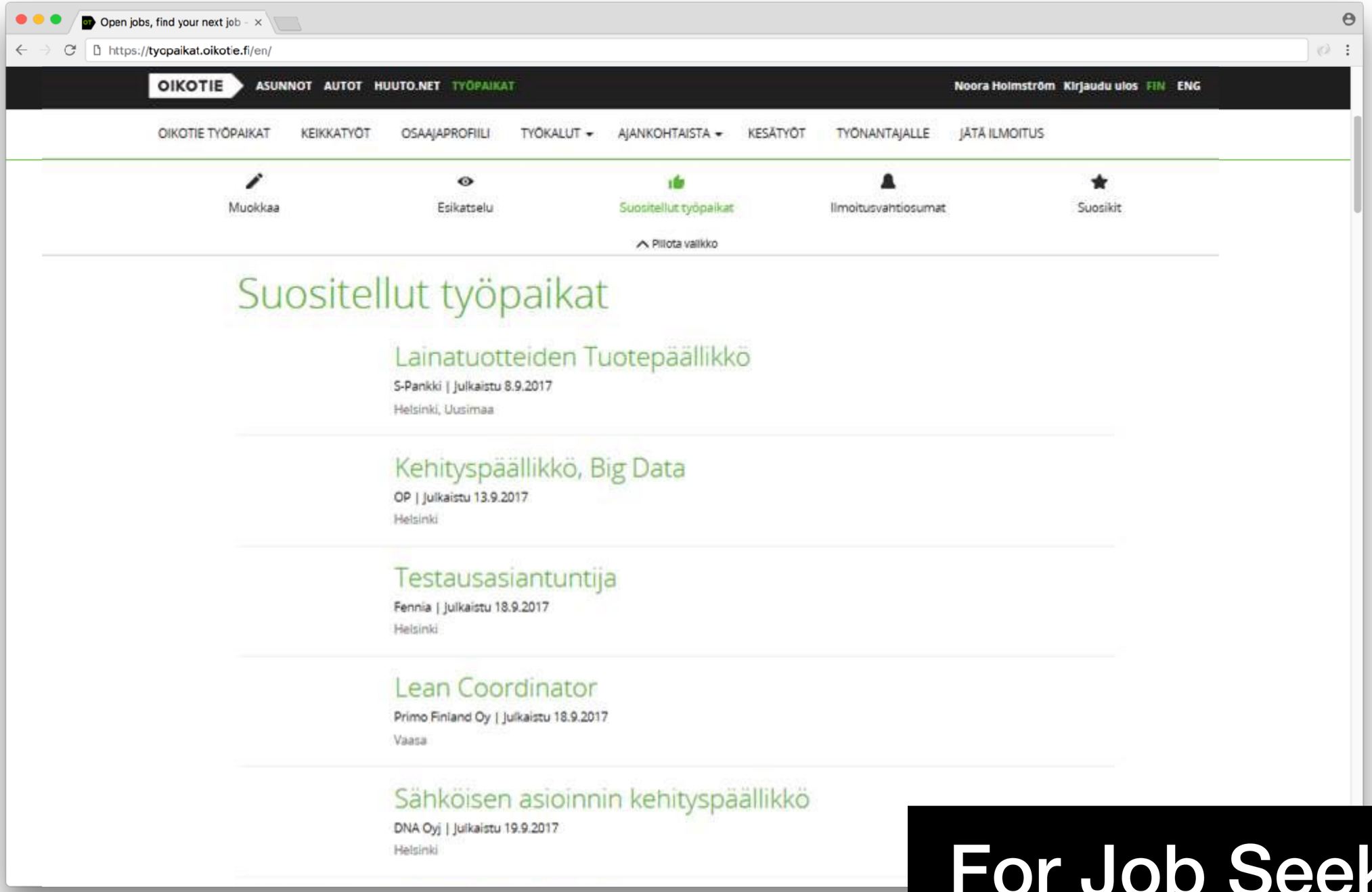
5. Questions

Intro

- Recruitment site **Oikotie Työpaikat** in Sanoma’s portfolio
- Recently introduced feature:
  automatic job and candidate discovery
- Recommendations powered by our algorithm



Oikotie Työpaikat



For Job Seekers

Open jobs, find your next job - ×

https://tyopaikat.oikotie.fi/en/

OIKOTIE ASUNNOT AUTOT HUUTO.NET TYÖPAIKAT My account Company Log out FIN ENG

RATKAISUT MY JOB ADS OSAAJAHAKU REC30 TUOTTEET TILASTOT

BALANCE 0 credits CART 0 products

Osaajahaku | Recommended profiles | Saved profiles

Recommended profiles

Here are some recommended profiles based on your job ads. You can search more from tens of thousands of profiles using the profile search. Try the profile search

SELECT THE AD FOR PROFILE MATCHING

Cloud Data Architect

Anne Analyytikko

Modified 12.2.2015

Analyytikko | Suomalainen yritys Oy

1.11.2012 –

Daniel Developer

Modified 28.11.2015

Company Oy

1.10.2016 – 31.8.2017

For Recruiters

tekniikka &talous

TEKNIIKAN HISTORIA

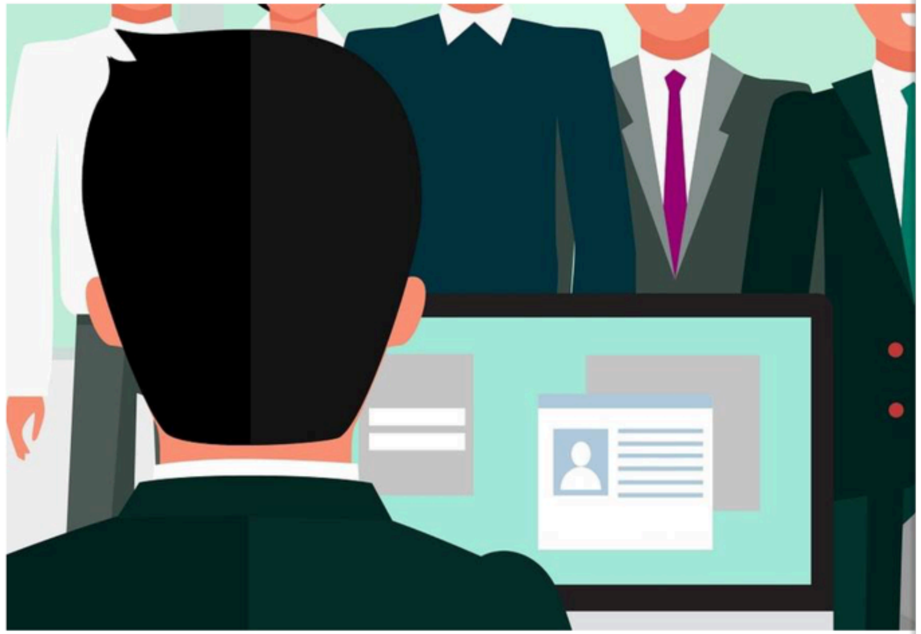
metalli-tekniikka

Tilaa leht

ETUSIVU | UUTISET | PUHEENVUOROT | KUMPPANIBLOGIT | OTA YHTEYTTÄ

REKRYTOINTI | Katja Ylinen 14.6. klo 14:58

Tekoäly tulee rekrytointiin - ”algoritmi suosittelee”



JAA ARTIKKELI

Oikotien työpaikkapalvelu on ottanut käyttöönsä uutta matching-teknologia… perustuu tekoälyyn.

Te

TYÖELÄMÄ

Keinoäly auttaa työnhakijoiden seulomisessa: "Algoritmi suosittelee avoimiin työpaikkoihin sopivia osaajaprofiileita"

15.6.2017 15:01 JURIDIIKKAFOKUS OMA DIGI ICT RAKENTAMINEN TALOUS JA POLITIIKKA

Oikotien työpaikkapalvelu on ottanut käyttöönsä uutta matching-teknologiaa, joka perustuu tekoälyyn.

OIKOTIEN työpaikkapalvelu on ottanut käyttöönsä uutta matching-teknologiaa, joka perustuu tekoälyyn.

Uusi palvelu perustuu Oikotien työpaikkailmoitusdatan ja Oikotielle syötettyjen osaajaprofiilien yhdistämiseen. Työhakijoiden tekemiä

Olet lukenut 0/5 maksutonta uutista

Te Katja Ylinen

Ota yhteyttä

Uusimmat

1 21:30 AUTO
Škodan menestys ärsyttää Volkswagenia - VW:n johto ja työntekijät haluavat hillitä sisarmerkki Škodan kilpailua

2 21:00 YHDYSVALTOJEN TALOUS
Trumpin budjettiesitys liikahti eteenpäin - edessä tulikoe

3 20:30 KAUPPA
Konkurssin jälkeen Seppälän työntekijät saivat palkkaturvasta nopeutetulla käsittelyllä 600 000 euroa – Anttilallekin maksettu jo 5 miljoonaa

4 20:00 DIGITALOUS
Opiskelijoiden vetämä ohjelmointitapahtuma ratkaisee Suomen hälyttävää koodaripulaa

The Algorithm In the News

The Challenge

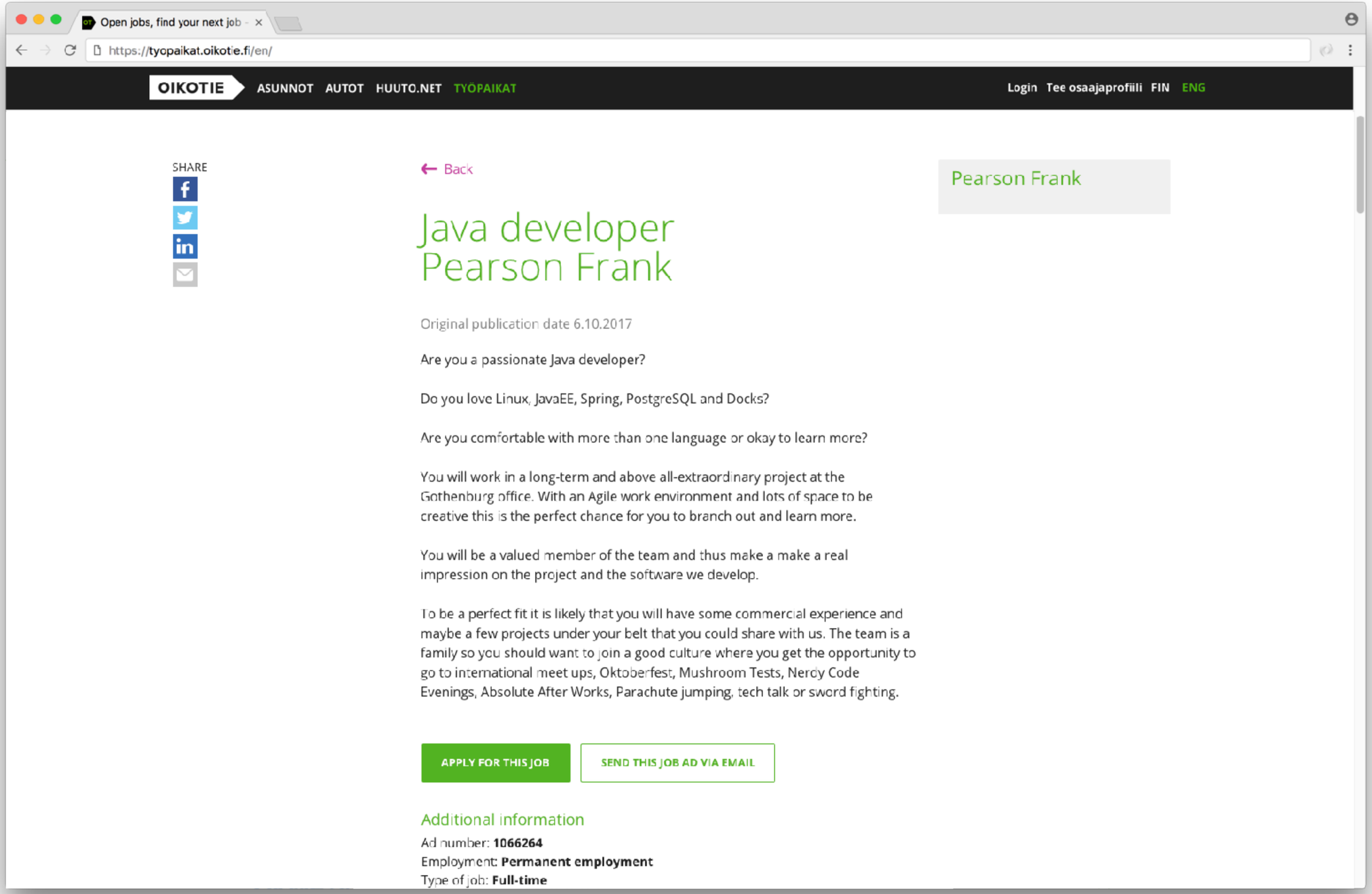
*“How might we **suggest***

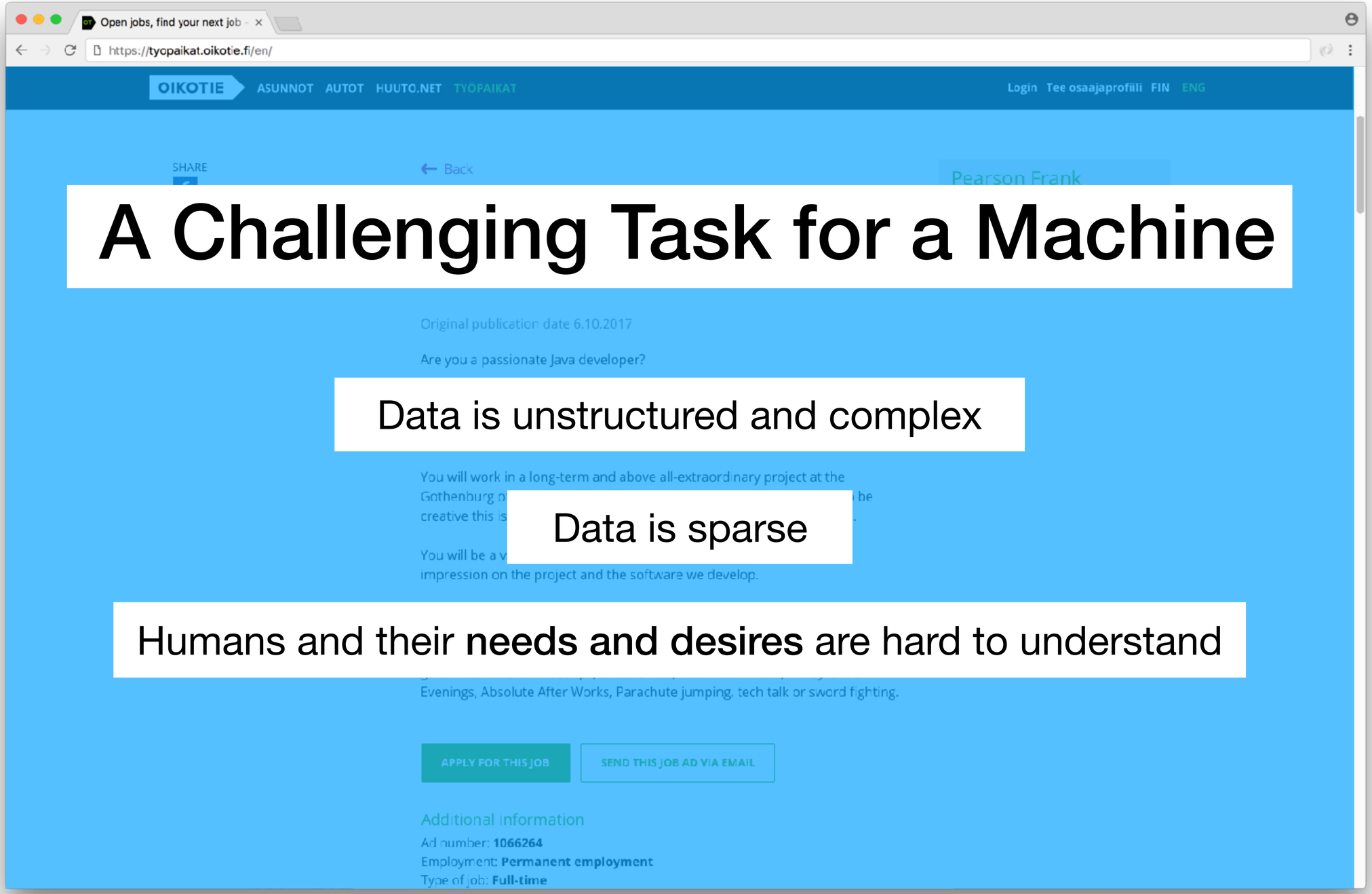
*the most **relevant positions** to a job seeker,*

*and vice versa **relevant candidates** to a recruiter,*

*to help both parties **find each other** more easily?”*

sanoma


Open jobs, find your next job -
https://tyopaikat.oikotie.fi/en/
OIKOTIE
ASUNNOT
AUTOT
HUUTO.NET
TYÖPAIKAT
Login
Tee osaajaprofiili
FIN
ENG
SHARE
Back
Java developer
Pearson Frank
Original publication date 6.10.2017
Are you a passionate Java developer?
Do you love Linux, JavaEE, Spring, PostgreSQL and Docks?
Are you comfortable with more than one language or okay to learn more?
You will work in a long-term and above all-extraordinary project at the Gothenburg office. With an Agile work environment and lots of space to be creative this is the perfect chance for you to branch out and learn more.
You will be a valued member of the team and thus make a make a real impression on the project and the software we develop.
To be a perfect fit it is likely that you will have some commercial experience and maybe a few projects under your belt that you could share with us. The team is a family so you should want to join a good culture where you get the opportunity to go to international meet ups, Oktoberfest, Mushroom Tests, Nerdy Code Evenings, Absolute After Works, Parachute jumping, tech talk or sword fighting.
APPLY FOR THIS JOB
SEND THIS JOB AD VIA EMAIL
Additional information
Ad number: 1066264
Employment: Permanent employment
Type of job: Full-time
Pearson Frank


Open jobs, find your next job
https://tyopaikat.oikotie.fi/en/
OIKOTIE
ASUNNOT
AUTOT
HUUTO.NET
TYOPAIKAT
Login
Tee osaajaprofiili
FIN
ENG
SHARE
Back
Pearson Frank
A Challenging Task for a Machine
Original publication date 6.10.2017
Are you a passionate Java developer?
Data is unstructured and complex
You will work in a long-term and above all-extraordinary project at the
Gothenburg o
be
creative this is
Data is sparse
You will be a v
impression on the project and the software we develop.
Humans and their needs and desires are hard to understand
Evenings, Absolute After Works, Parachute jumping, tech talk or sword fighting.
APPLY FOR THIS JOB
SEND THIS JOB AD VIA EMAIL
Additional information
Ad number: 1066264
Employment: Permanent employment
Type of job: Full-time

Approaches

A Small Deep Dive into Machine Learning

sanoma

Approaches

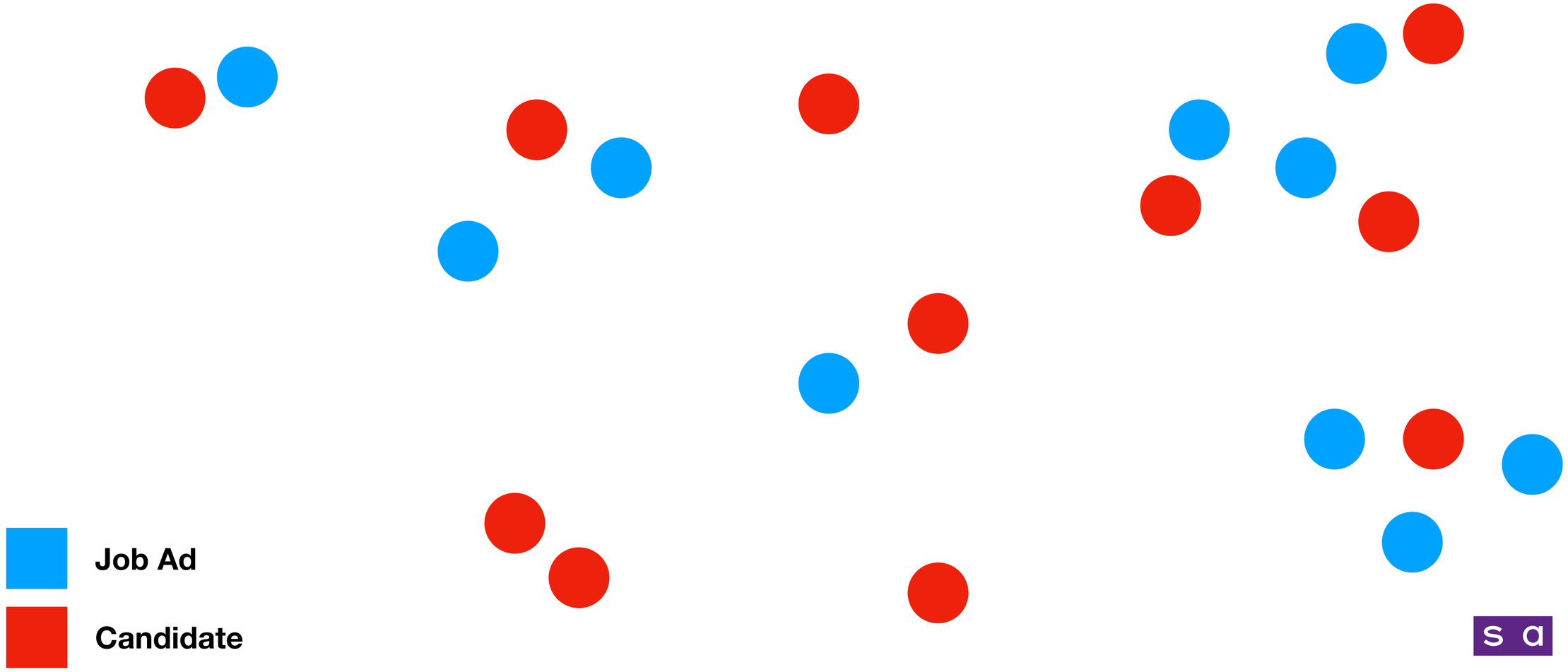
Vector Space Models

- Relevance = distance
- Candidates and Jobs are mapped into a space
- Find closest matches with a distance metric
- Space is high-dimensional

s a
o m a
n

Approaches

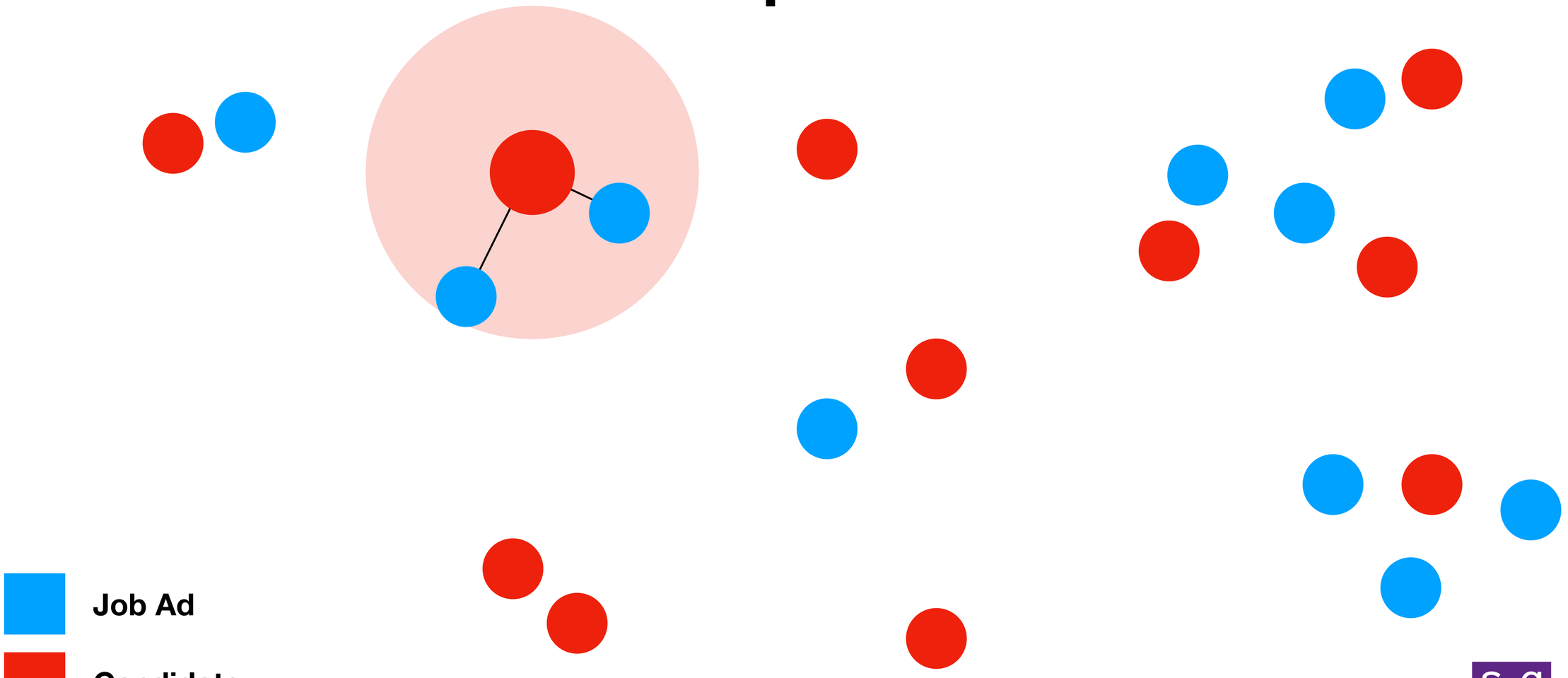
Vector Space Models



sanoma

Approaches

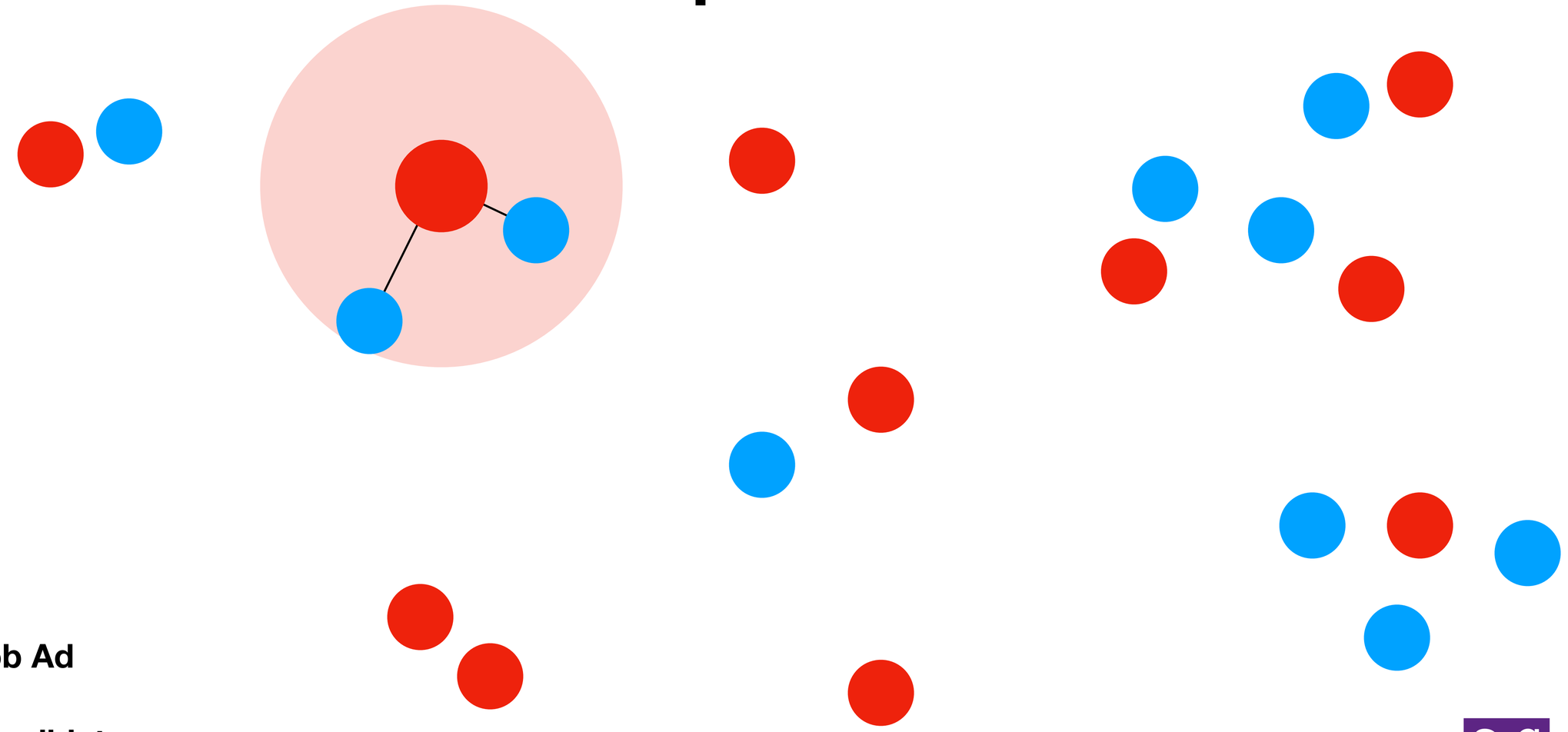
Vector Space Models



sanoma

Approaches

Vector Space Models



Job Ad

Candidate

And now just imagine this with a 1000 dimensions instead.

sanoma

Approaches

Constructing the Vector Space

- Basic search techniques on metadata
- Natural Language Processing
  - Keyword matching
  - **Topic Modeling with Latent Dirichlet Allocation (LDA)**
  - **Continuous Distributed Representations**
- Behavioral Analysis
  - **Collaborative Filtering**
- Many other techniques

sa
noma

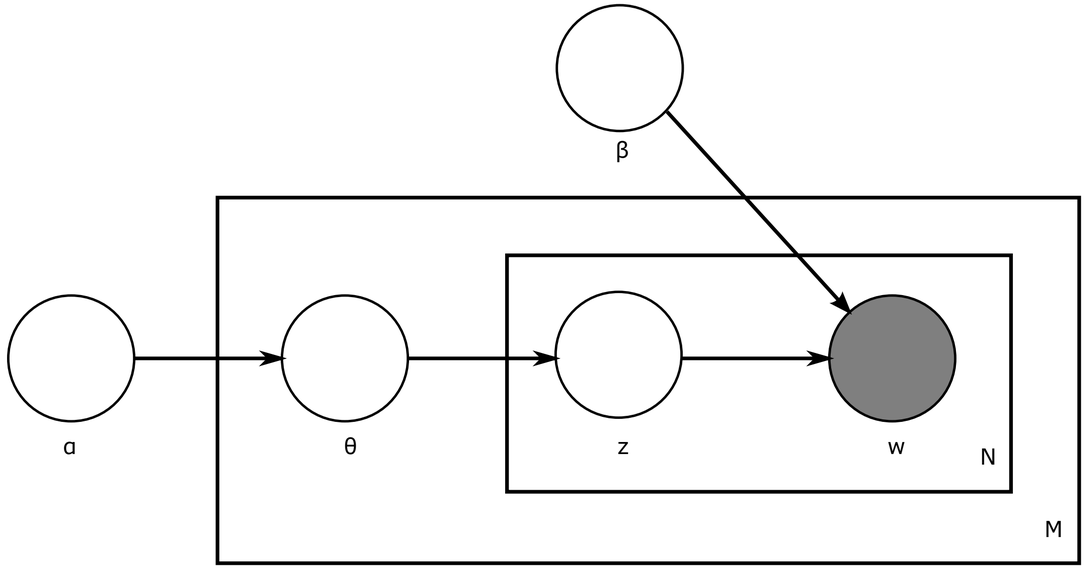
Approaches

Topic Modeling with Latent Dirichlet Allocation (LDA)

- **Unsupervised Method** by [Blei et al., 2003]
- Assumes process of creating documents:

  First: choose a few topics, then:

  1. Pick a topic
  2. Pick a word associated with the topic
  3. Repeat until we have enough words



s a
o m a
n

Approaches

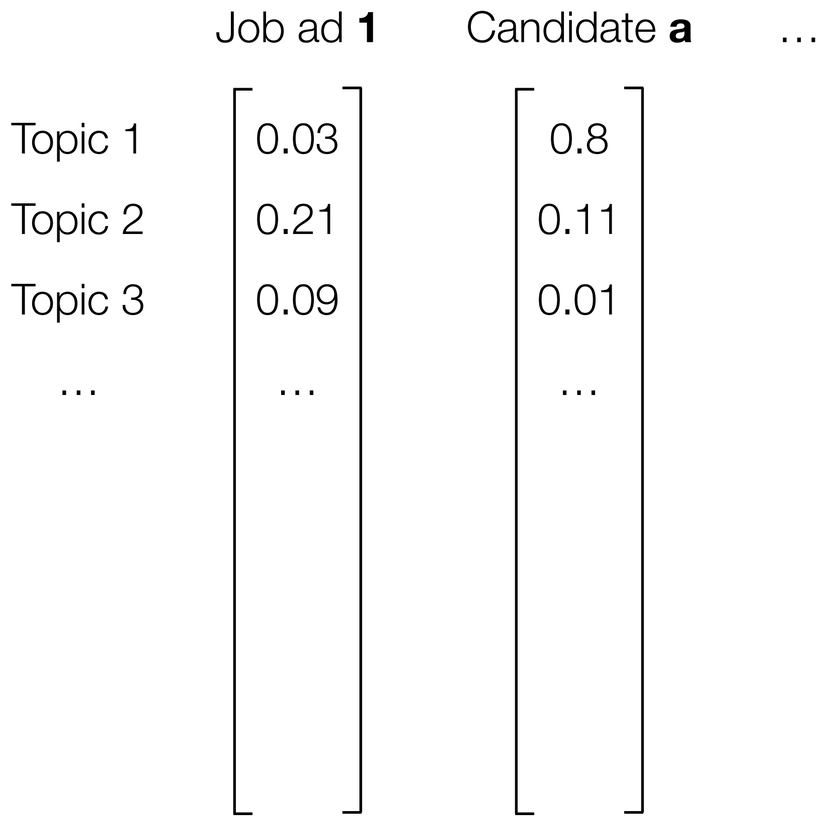
Topic Modeling with Latent Dirichlet Allocation (LDA)

- Learning using Bayesian Inference
- Yields probabilities, how likely

  A. each word occurs with each topic

  B. each topic occurs in each document

- For new documents:
  infer its association with the topics

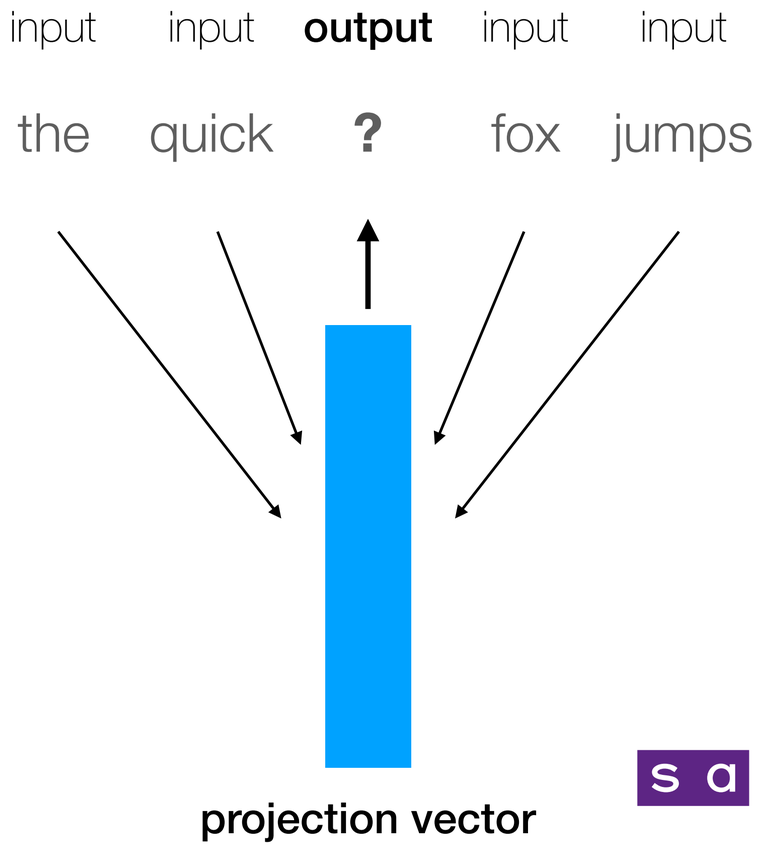


sanoma

Approaches

Continuous Distributed Representations

- Commonly known as “word2vec”
  [Mikolov, 2013], roots in [Hinton, 1986]
- Neural Network based
- Learn word representations by predicting in which context a word appears
- Needs huge amounts of data to work (e.g. all of wikipedia)



s a
o m a
n

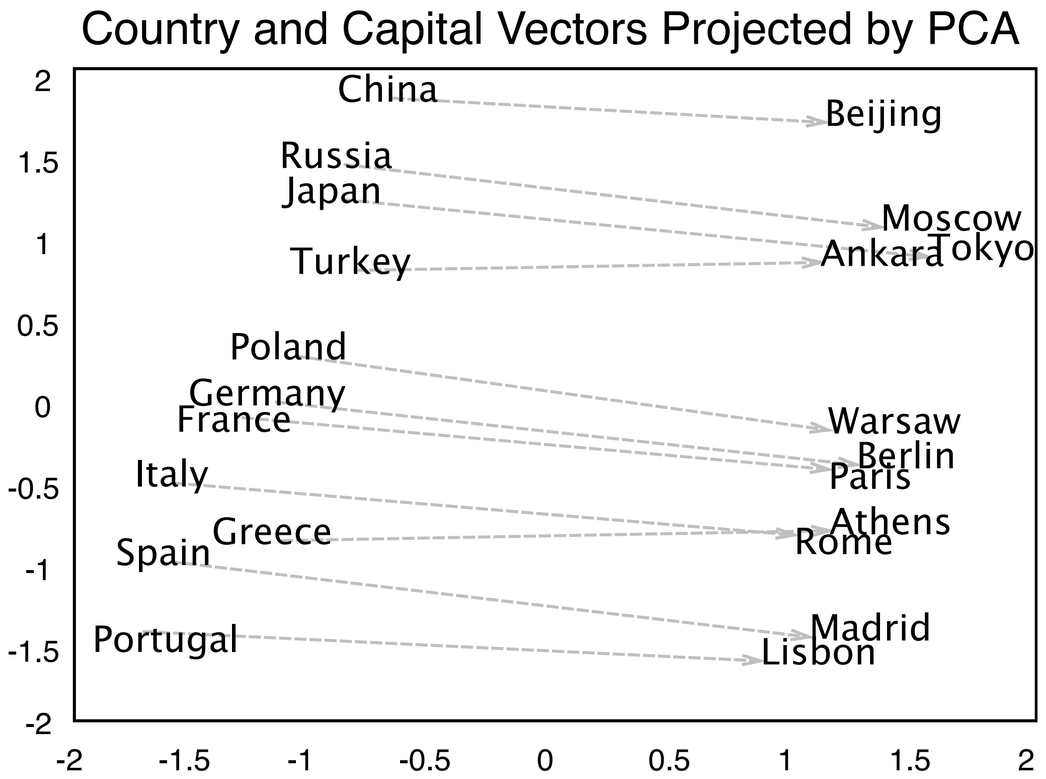
Approaches

Continuous Distributed Representations

- Captures surprising semantic properties:

```
v(Berlin) - v(Germany) + v(France) = v(Paris)
```

- This allows e.g. for analogy queries
- Several techniques to extend “word vectors” to “document vectors”
- Allows us again to map job ads and candidate profiles to the vector space



Approaches

Behavioral Analysis: Collaborative Filtering

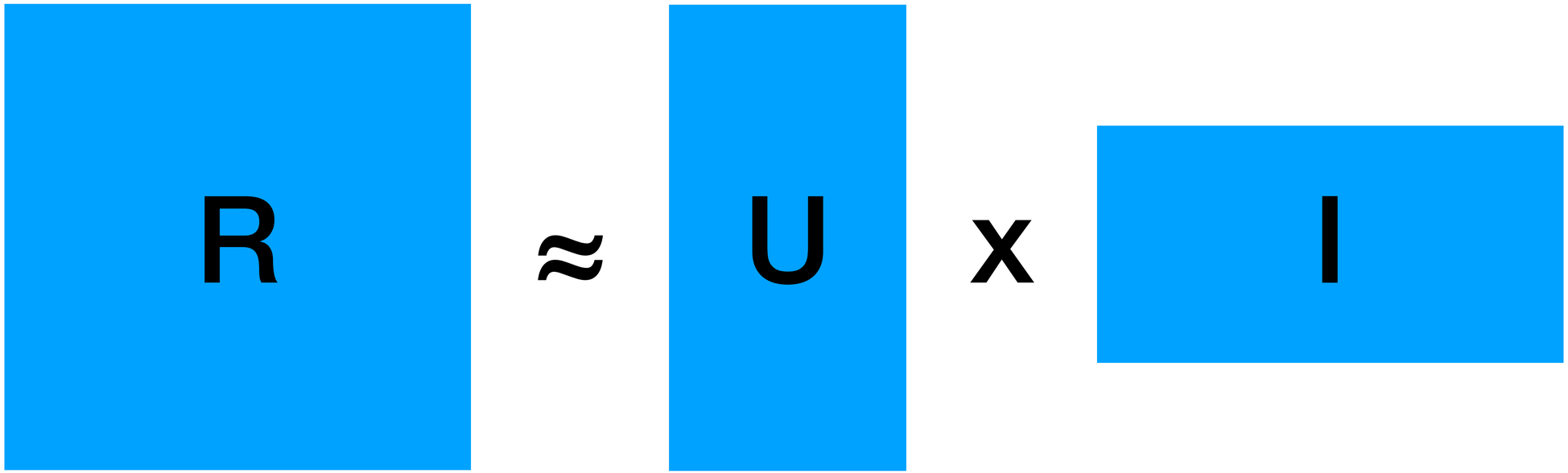
	Orange is the new Black	Stranger Things	Narcos	House of Cards	Daredevil
JOHN	5		1	2	2
LUCY	1	2	5	5	5
DIANE	4	5	3	3	
YOU	2	3	?	5	4

Users x Items matrix **R**

sanoma

Approaches

Matrix Factorization



Users x Items **Users x Factors** **Factors x Items**

sanoma

Matching in Oikotie Työpaikat

- We use a mixture of similar and additional approaches
- An API serves the results, e.g. like this:

```
{
  “results”:[
    {“id”: 1050236, “confidence”: 0.92},
    {“id”: 2572425, “confidence”: 0.81},
    {“id”: 1235285, “confidence”: 0.73},
    {“id”: 3413478, “confidence”: 0.69}
  ]
}
```

sanoma

Evaluation

- Qualitative: Interviews, user testing, surveys
- Quantitative:
  - Verify with implicit feedback (job ad visits, “apply”-clicks)
  - Implement explicit feedback mechanism
  - Test with rating experiment
  - Can be framed as prediction or ranking problem

Learnings and Takeaways

Plan well and follow a process (e.g. double diamond)

- User need must be understood (the problem scope)
- Prototype: fail and learn (the solution scope)
- Productization according to software quality standards takes 3x more time than you think
- Don’t forgot maintenance, privacy assessments, handovers

Capture your problem with a metric (or several ones)

- Measure from the start (validates results and helps communicating them)
- Update if needs change

sanoma

Gracias

Time for Questions

Clemens Westrup
firstname.lastname@sanoma.com