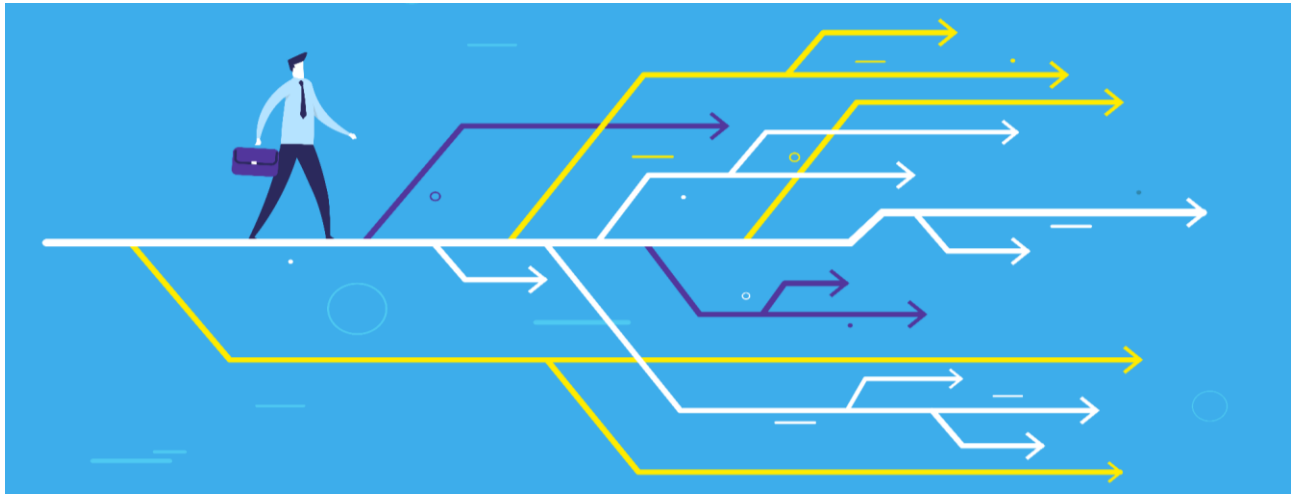
EXPLORING TF-IDF ENCODING FOR COMPREHENSIVE INDUSTRY PARTNERS' SELECTION

Presenter: Yuri Campbell

Co-authors: Karl Trela, Friedrich Dornbusch

Fraunhofer IMW, Leipzig



September 11th 2018

Global Tech Mining Conference 2018, Leiden

Scientific Session: Methods in Technology Management

© Fraunhofer IMW

Fraunhofer IMW

THE FRAUNHOFER-GESELLSCHAFT

EUROPE'S LARGEST ORGANIZATION FOR APPLIED RESEARCH

- 72 institutes and facilities
- More than 25,000 employees
- 2.3 billion Euro research budget; with almost 2 billion Euro in contractual research

© Fraunhofer IMW

Fraunhofer IMW

Willkommen im Fraunho
Sicher | https://info.fraunhofer.de/Seiten/default.aspx

Fraunhofer

New Search

CRL-Principle

FAQ

SME-Match

research institute

technology, use-case, field of expertise

similar companies

Search

About us

Windows durchsuchen
14:38
10.07.2017

Willkommen im Fraunho
Sicher | https://info.fraunhofer.de/Seiten/default.aspx

Fraunhofer

New Search
CRL-Principle
FAQ

SME-Match

Fraunhofer IOF
Fraunhofer IBP
Fraunhofer ESK
Fraunhofer IAF
Fraunhofer EMI
Fraunhofer ICT
Fraunhofer LBF
Fraunhofer IOSB
Fraunhofer IPA

Search

About us

Windows durchsuchen
14:38
10.07.2017

Willkommen im Fraunhof
Sicher | https://info.fraunhofer.de/Seiten/default.aspx

Fraunhofer

SME-Match

New Search
CRL-Principle
FAQ

Institute: Fraunhofer IWS
Search: LED technology
Similar: Silicia GmbH

CRL	Name	City	State	Employees	Turnover	Description
8.9	electronic service willms gmbH & Co. KG	Stolberg (Rhld.)	Nordrhein-Westfalen	[illegible]	[illegible]	[illegible]
8.9	Western Digital Deutschland GmbH	Aschheim	Bayern	[illegible]	[illegible]	[illegible]
8.8	JDI Europe GmbH	München	Bayern	[illegible]	[illegible]	[illegible]
8.8	Gemalto M2M GmbH	München	Bayern	[illegible]	[illegible]	[illegible]
8.7	Kyocera Fineceramics GmbH	Esslingen am Neckar	Baden-Württemberg	[illegible]	[illegible]	[illegible]
8.2	ODU GmbH & Co. KG	Mühldorf a. Inn	Bayern	[illegible]	[illegible]	[illegible]
8.2	Kyocera Display Europe GmbH	Dietzenbach	Hessen	[illegible]	[illegible]	[illegible]
7.8	Airbus DS Airborne Solutions GmbH	Bremen	Bremen	[illegible]	[illegible]	[illegible]
7.7	Bourns Sensors GmbH	Taufkirchen	Bayern	[illegible]	[illegible]	[illegible]
6.3	PAS Deutschland GmbH	Neuruppin	Brandenburg	[illegible]	[illegible]	[illegible]
6.2	Humax Digital GmbH	Unterföhring	Bayern	[illegible]	[illegible]	[illegible]
6.1	cms electronics gmbh	Klagenfurt	Kärnten	[illegible]	[illegible]	[illegible]
6.0	FORTEC Elektronik Aktiengesellschaft	Landsberg am Lech	Bayern	[illegible]	[illegible]	[illegible]
6.0	Turck Duotec GmbH	Halver	Nordrhein-Westfalen	[illegible]	[illegible]	[illegible]
6.0	Littlebit Technology GmbH	Alzenau	Bayern	[illegible]	[illegible]	[illegible]
5.8	TECTRON GmbH Worbis - Entwicklung unc	Leinefelde-Worbis	Thüringen	[illegible]	[illegible]	[illegible]
5.2	Unimicron Germany GmbH	Geldern	Nordrhein-Westfalen	[illegible]	[illegible]	[illegible]
4.4	cab Produkttechnik Gesellschaft für Comp	Karlsruhe	Baden-Württemberg	[illegible]	[illegible]	[illegible]
4.4	FCI Deutschland GmbH	Oberursel (Taunus)	Hessen	[illegible]	[illegible]	[illegible]
4.3	Gebauer & Griller Metallwerk GmbH	Wien	Wien	[illegible]	[illegible]	[illegible]

About us

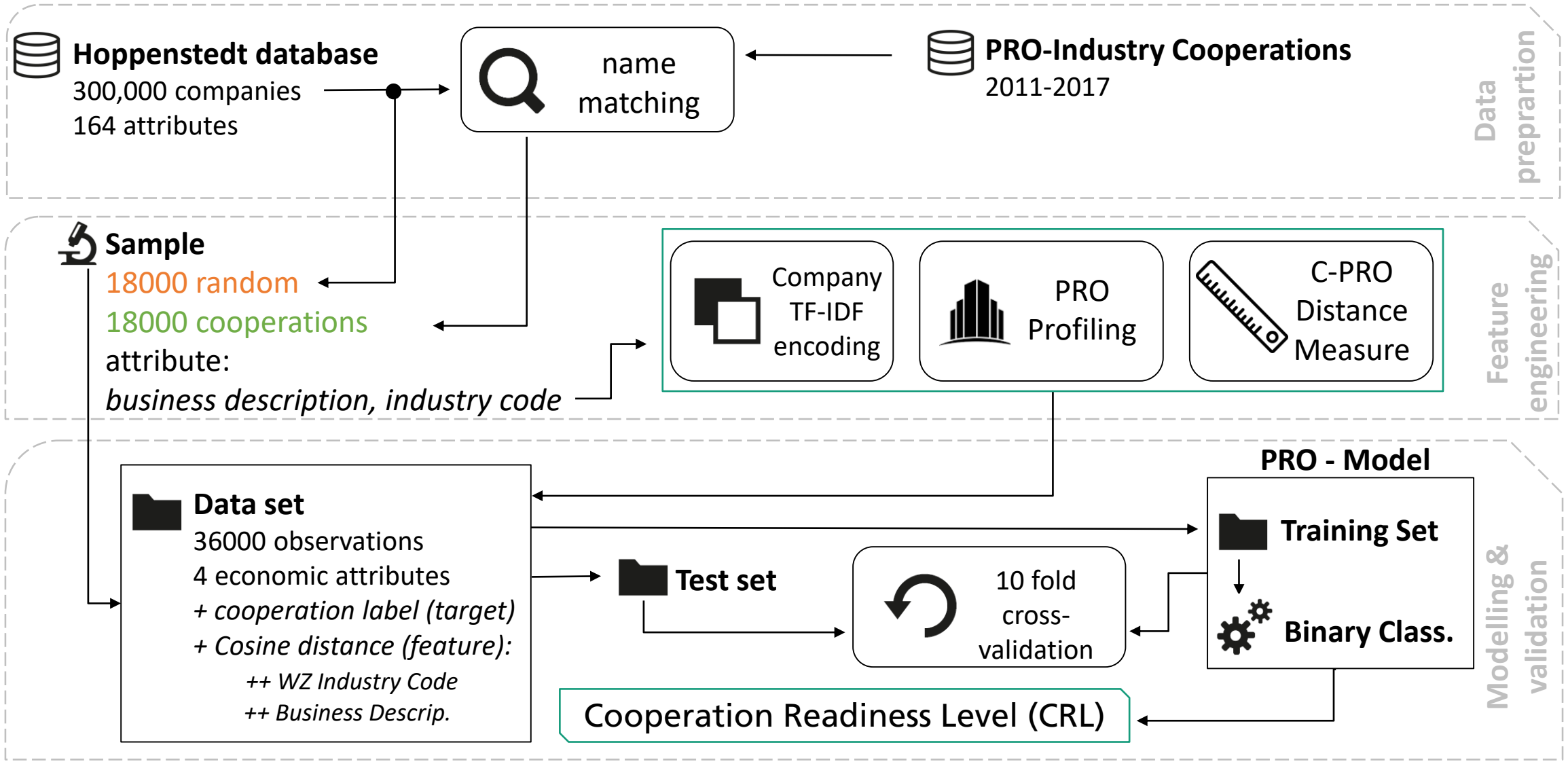
Distance
- ○ 0km
- ○ 1-5km
- ○ 6-20km
- ○ 21-50km
- ○ 51-100km
- ○ >100km

Employees
- ○ 1-20
- ○ 21-50
- ○ 51-100
- ○ 101-250
- ○ >250

Turnover
- ○ <2.5M€
- ○ 2.5-5M€
- ○ 5-12.5M€
- ○ 12.5-25M€
- ○ >25M€

Windows durchsuchen
14:38
10.07.2017

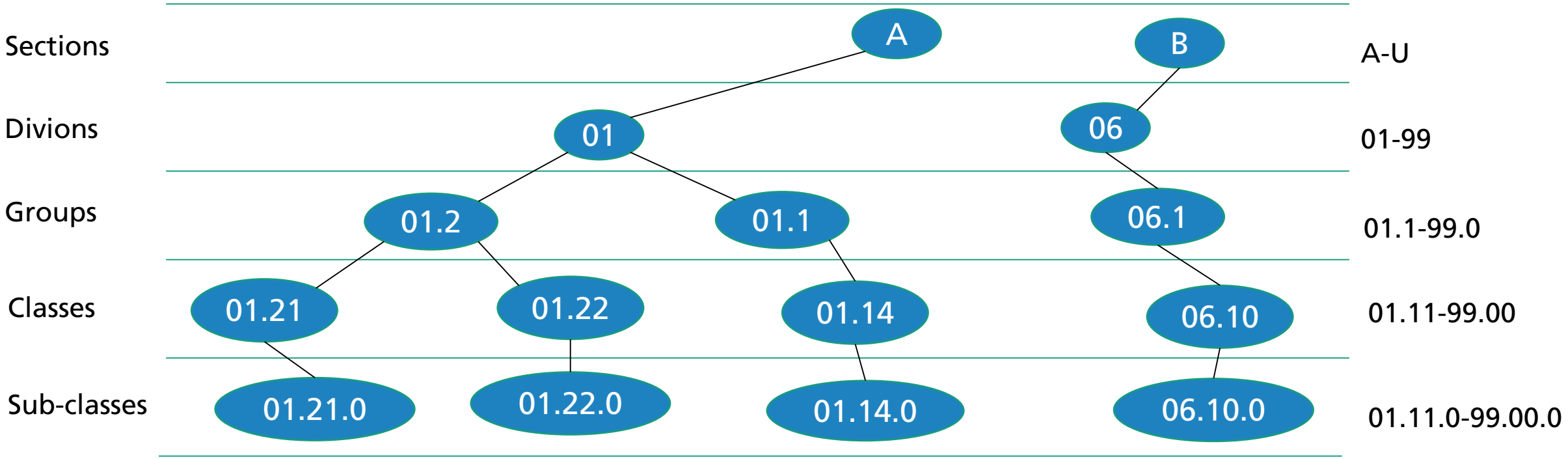
Statistical model



© Fraunhofer IMW

Fraunhofer IMW

WZ Industry Code



A-01.21.0
Growing of Grapes

A-01.22.0
Growing of Tropical Fruits

A-01.14.0
Growing of Sugar Cane

B-06.10.0
Extraction of Crude Petroleum

© Fraunhofer IMW
Fraunhofer IMW

WZ Industry Code Encoding

	WZ Code	Generated Text	TF-IDF Encoding
Growing of Grapes	A-01.21.0	A A01 A012 A0121 A01210	$\vec{w}_{grapes}$
Growing of Tropical Fruits	A-01.22.0	A A01 A012 A0122 A01220	$\vec{w}_{trop.fruits}$
Growing of Sugar Cane	A-01.14.0	A A01 A011 A0114 A01140	$\vec{w}_{sugarcane}$
Extraction of Crude Petroleum	B-06.10.0	B B06 B061 B0610 B06100	$\vec{w}_{petroleum}$

$$cosdis(\vec{w}_{grapes}, \vec{w}_{trop.fruits}) < cosdis(\vec{w}_{grapes}, \vec{w}_{sugarcane}) < cosdis(\vec{w}_{grapes}, \vec{w}_{petroleum}) = 1$$

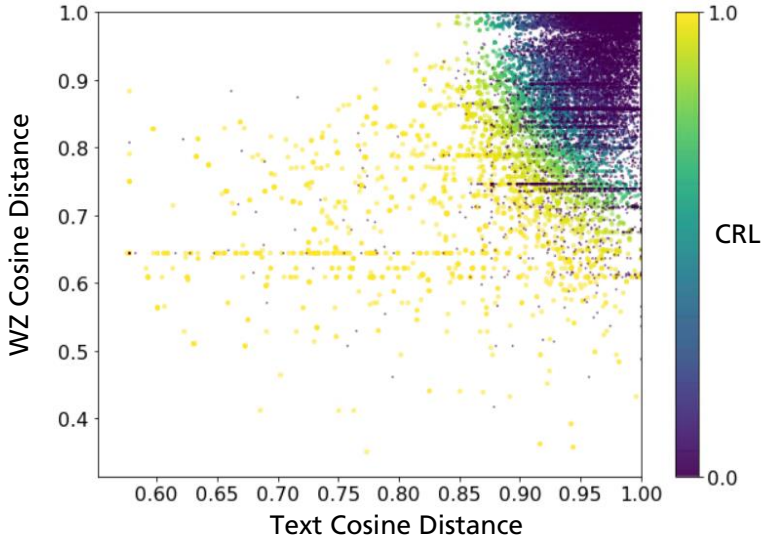
© Fraunhofer IMW

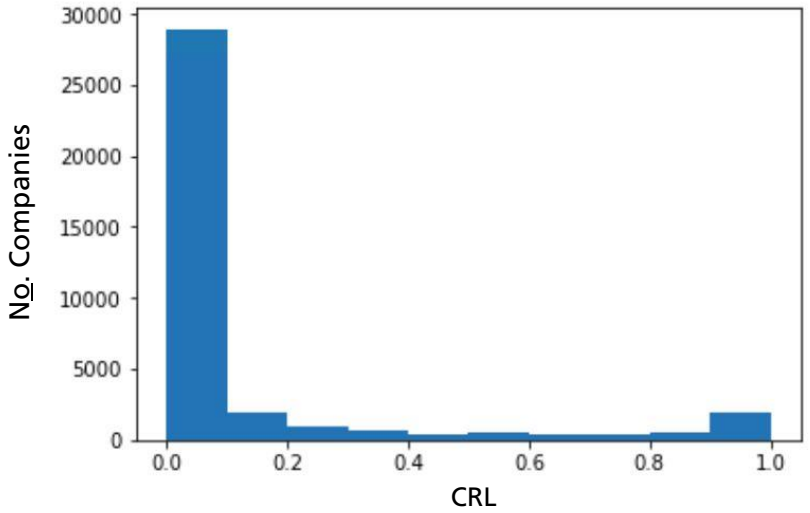
Results

Unit 133

23 partners

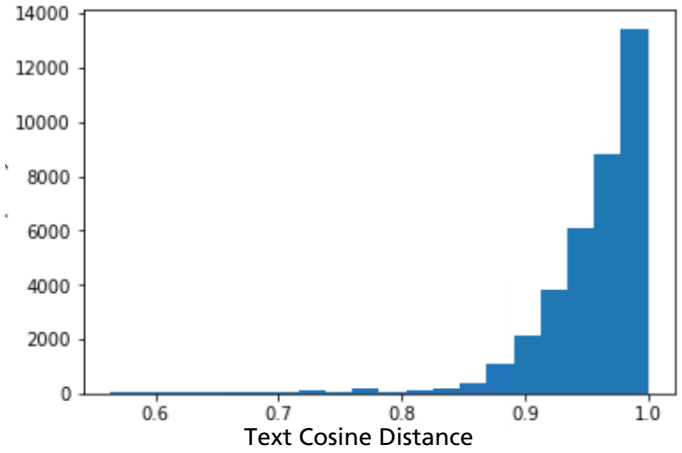
Cooperation Readiness Level (CRL)



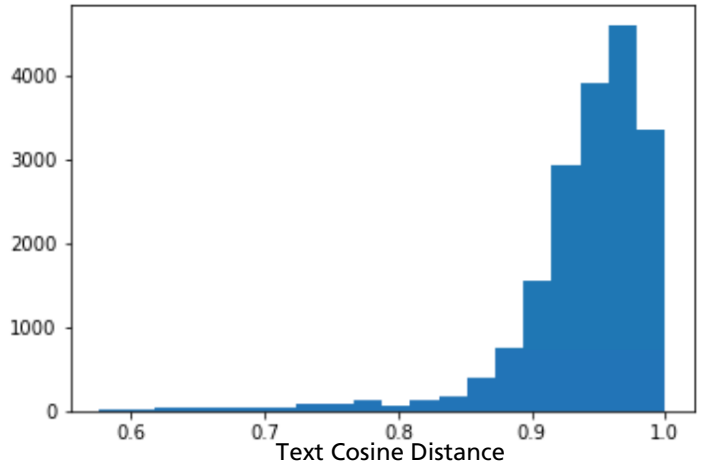


Text Cosine Distance Histograms

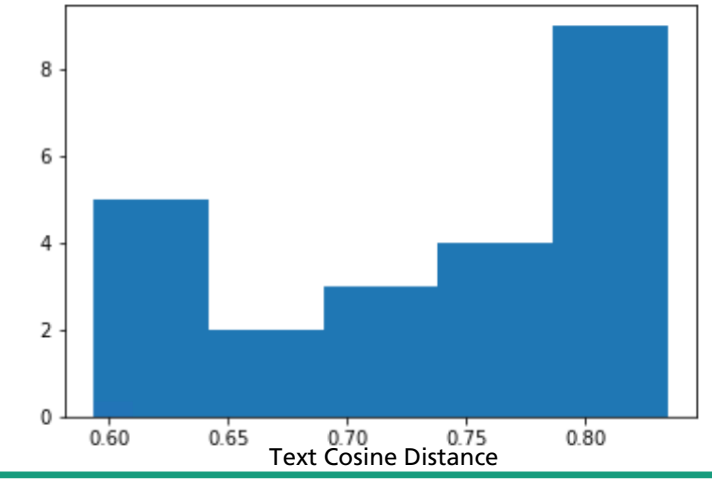
All Companies



Fraunhofer Partners



Unit 133 Partners



© Fraunhofer IMW

Fraunhofer IMW

Final Remarks

- TF-IDF information encoding technique beyond text
  - Explore code structure
  - Classification Systems
- Indirect approach:
  - Similarity is measured among players of the same type
- Recommendation list:
  - Average ROC AUC score of:
    - *0.83*, with std. deviation of *0.06,* w/ strong data separation
    - 0.90, with std. deviation of *0.06,* w/ soft data separation
  - Probability of falsely ranking random chosen pair
- TO DO’s:
  - Include more features (Patent information, public funded projects, semantic measures)
  - Test score measures & compare to other methods

© Fraunhofer IMW

Term frequency–inverse document frequency (TF-IDF)

SWAREFLEX GmbH

The company is manufacturing and selling electronic road safety products with LEDs (including glass reflectors for road markings).

company

manufacturing road reflector

selling safety product

electronic LED

markings

road including

glass

$$\mathrm{TFIDF}_{t,SWAREFLEX} = tf_{t,SWAREFLEX} \cdot \log \frac{N}{df_t}$$

$$\mathrm{TFIDF}_{reflector,SWAREFLEX} = tf_{reflector,SWAREFLEX} \cdot \log \frac{N}{df_{reflector}}$$

$$\mathrm{TFIDF}_{reflector,SWAREFLEX} = \frac{1}{13} \cdot \log \frac{36000}{3} = 0.24$$

$$\vec{t}_{SWAREFLEX} = \begin{bmatrix} 0.27 \\ 0.24 \\ 0.12 \\ \vdots \\ \mathrm{TFIDF}_{V,d1} \end{bmatrix}$$

© Fraunhofer IMW

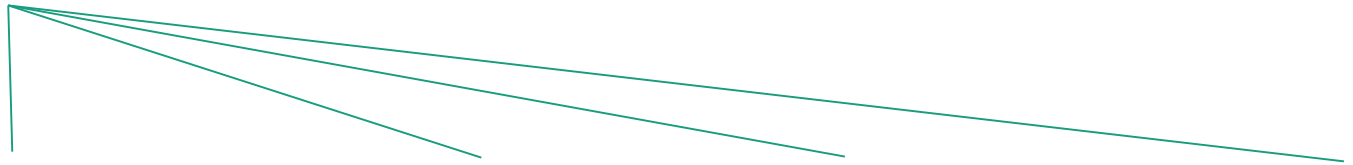
PRO Profile

Fraunhofer IOF

Fraunhofer Institute for Applied Optics and Precision Engineering IOF

https://www.iof.fraunhofer.de/en/business-fields/optical-components-and-systems/led-illumination-systems.html

PRO



SWAREFLEX GmbH | Company 1 | Company 2 | ⋯ | Company n

Partners

$\vec{t}_{SWAREFLEX}$ $\vec{t}_1$ $\vec{t}_2$ $\vec{t}_n$

$\vec{w}_{SWAREFLEX}$ $\vec{w}_1$ $\vec{w}_2$ $\vec{w}_n$

$$T_{IOF} = \sum_i \vec{t}_i \qquad W_{IOF} = \sum_i \vec{w}_i$$

Profile

$$d_{IOF,SWAREFLEX} = 1 - \cos(T_{IOF}, \vec{t}_{SWAREFLEX}) \qquad d_{IOF,SWAREFLEX} = 1 - \cos(T_{IOF}, \vec{w}_{SWAREFLEX})$$

Feature

Text | WZ Code

© Fraunhofer IMW

Fraunhofer IMW