



# Patentsuche und –überwachung mit Methoden der künstlichen Intelligenz in Theorie und Praxis

Teil 1: Maschinelles Lernen für Patentklassifizierung

Kornél Markó, Averbis GmbH



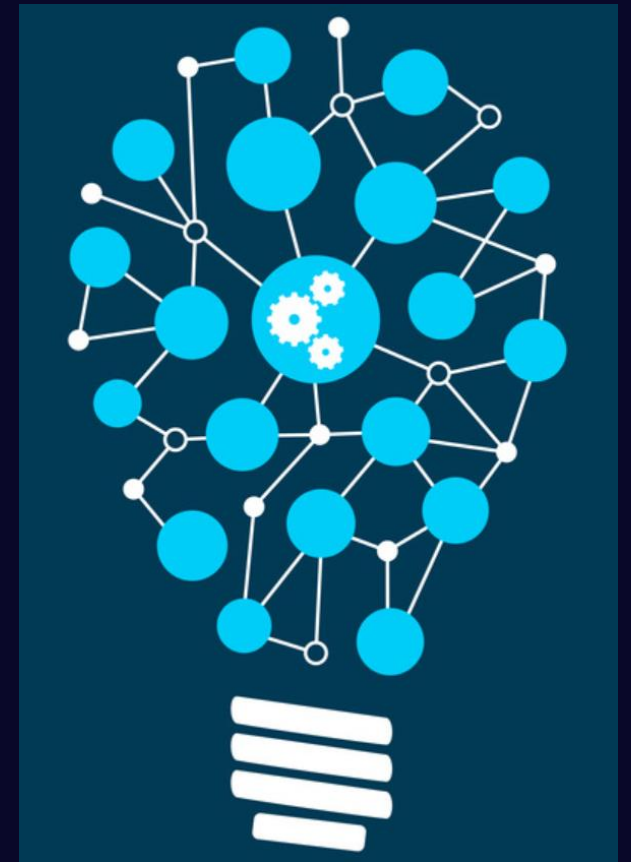
**Fokus:** Automatisierung von kognitiven  
Routinetätigkeiten

**Technologien:** Text Mining, Machine Learning, Semantik

**Produkte:** Information Discovery,  
Patent Monitor, Health Discovery

**Märkte:** Healthcare, Pharma, Biotech, Automotive,  
Bibliotheken, ...

**Gegründet:** 2007 in Freiburg



# Über Averbis

- Portfolioanalyse:** Bewerten von Stärken und Schwächen eines Patentportfolios
- Mitbewerberanalyse:** Identifizierung technologischer und strategischer Informationen über aktuelle und zukünftige Wettbewerber
- Landscaping:** Veranschaulichung der Patentsituation zu einer bestimmten Technologie innerhalb eines bestimmten Landes/Region/Global
- Innovation:** Identifizierung von Markttrends und neuen Technologien
- Freedom-to-Operate:** Produktentwicklung, ohne das geistige Eigentum anderer zu verletzen

## Hintergrund Patentanalysen

## Bei unzähligen neuen Patentanmeldungen besteht die Notwendigkeit...



... keine Zeit mehr mit endloser Routinetätigkeit zu verbringen



... die Genauigkeit bei der Patentidentifikation massiv zu verbessern



... IP-Experten zu befähigen, sich auf die interessante Arbeit zu konzentrieren



... IP-Experten zu befähigen, datengetriebene Entscheidungen zu treffen

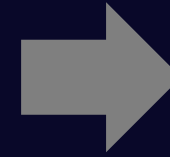
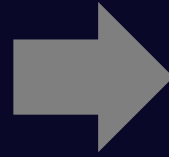
# Die Herausforderung

- ... ist eine auf *Maschinellem Lernen* basierende Software für die automatische Patentklassifikation
- ... analysiert eine große Anzahl von Patenten (und NPL), schätzt deren Relevanz mit hoher Genauigkeit und klassifiziert sie automatisch in *nutzerspezifische* Kategorien
- ... lernt kontinuierlich von und imitiert das Verhalten von IP-Experten



## Traditioneller Ansatz

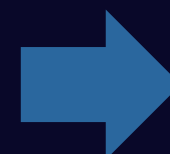
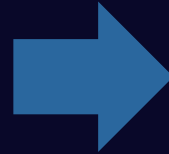
Daten  
Programm



Ergebnis

## Maschinelles Lernen

Daten  
Beispiele



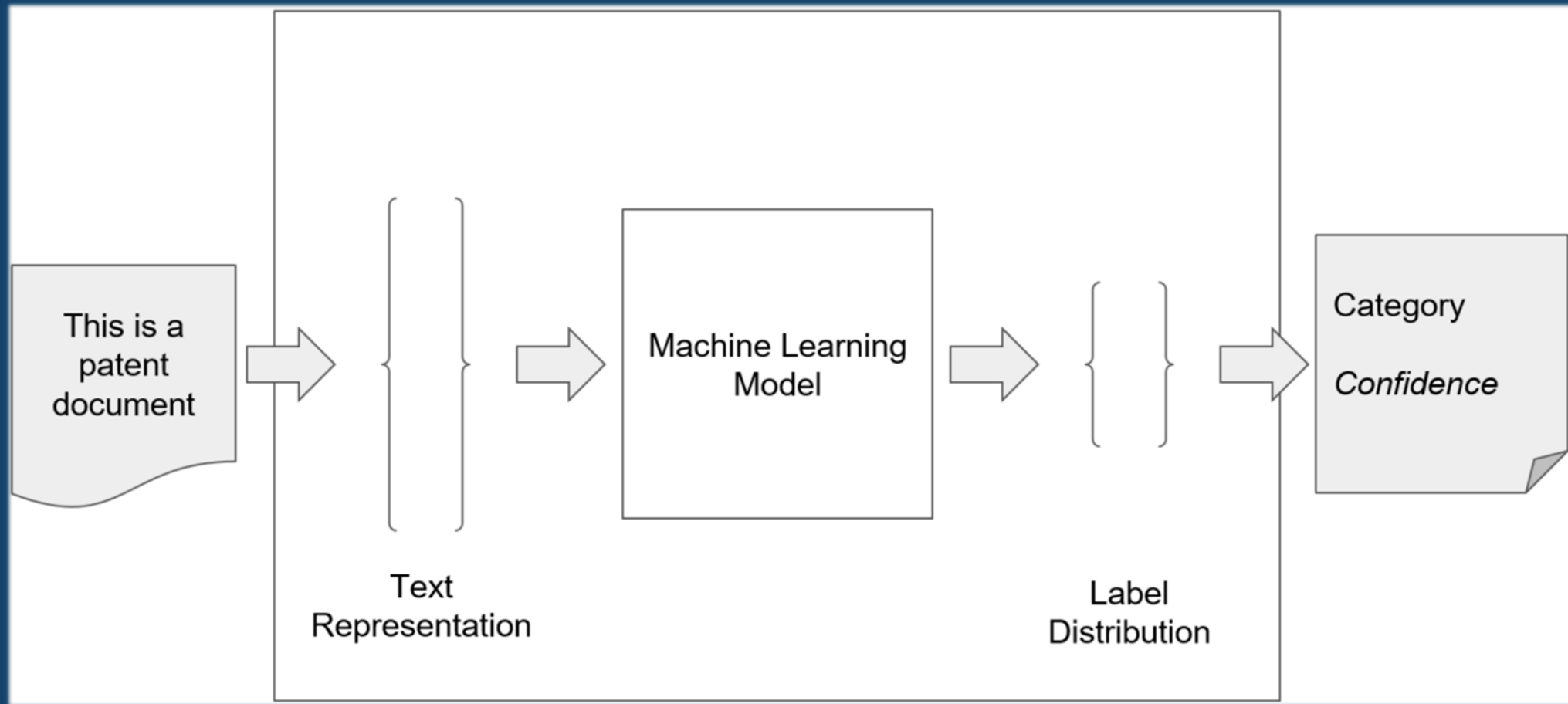
Programm

**Was ist Maschinelles Lernen?**

- Naïve Bayes
- Decision Trees
- *Support Vector Machines*
- Neural Networks
- *Convolutional Neural Networks*
- Language Modeling

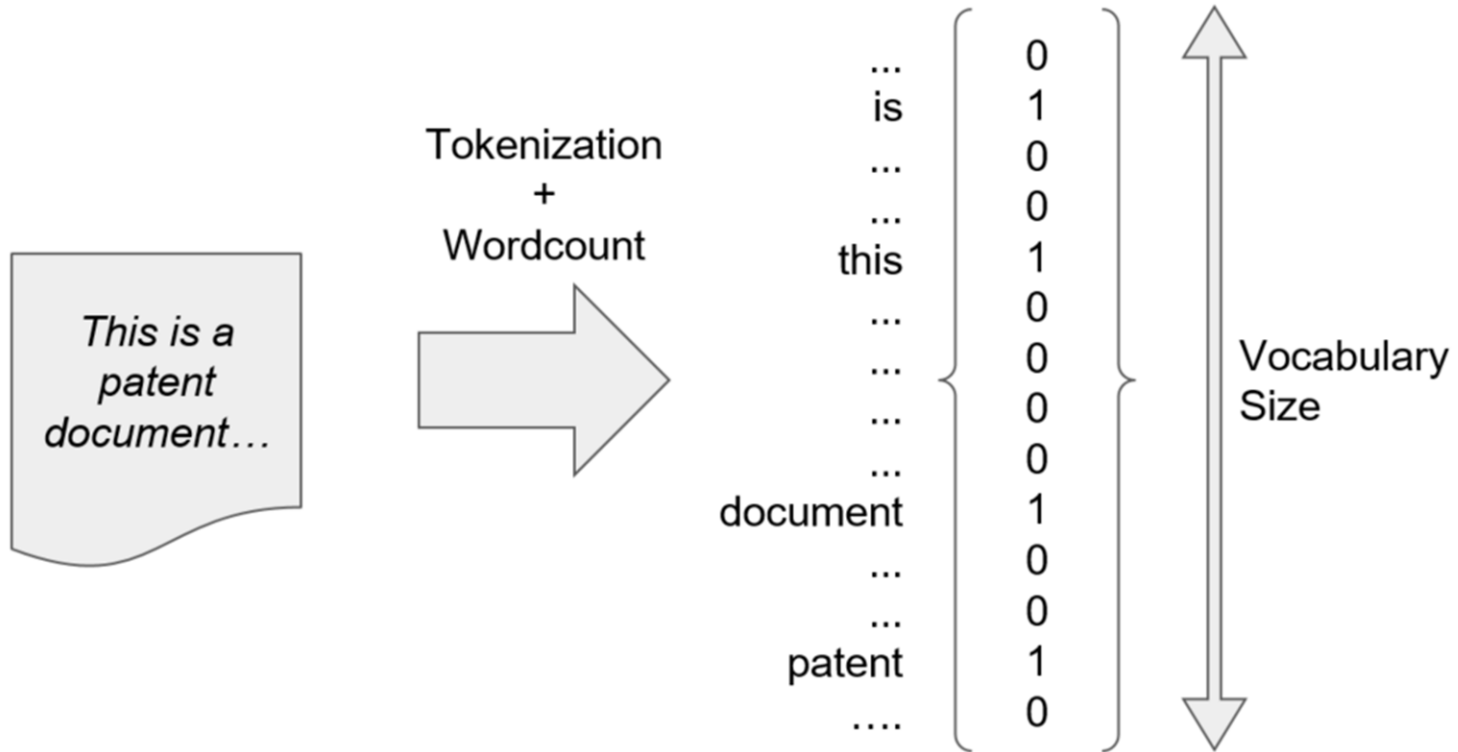
...

**Modelle für Maschinelles Lernen**



# Maschinelles Lernen: Workflow



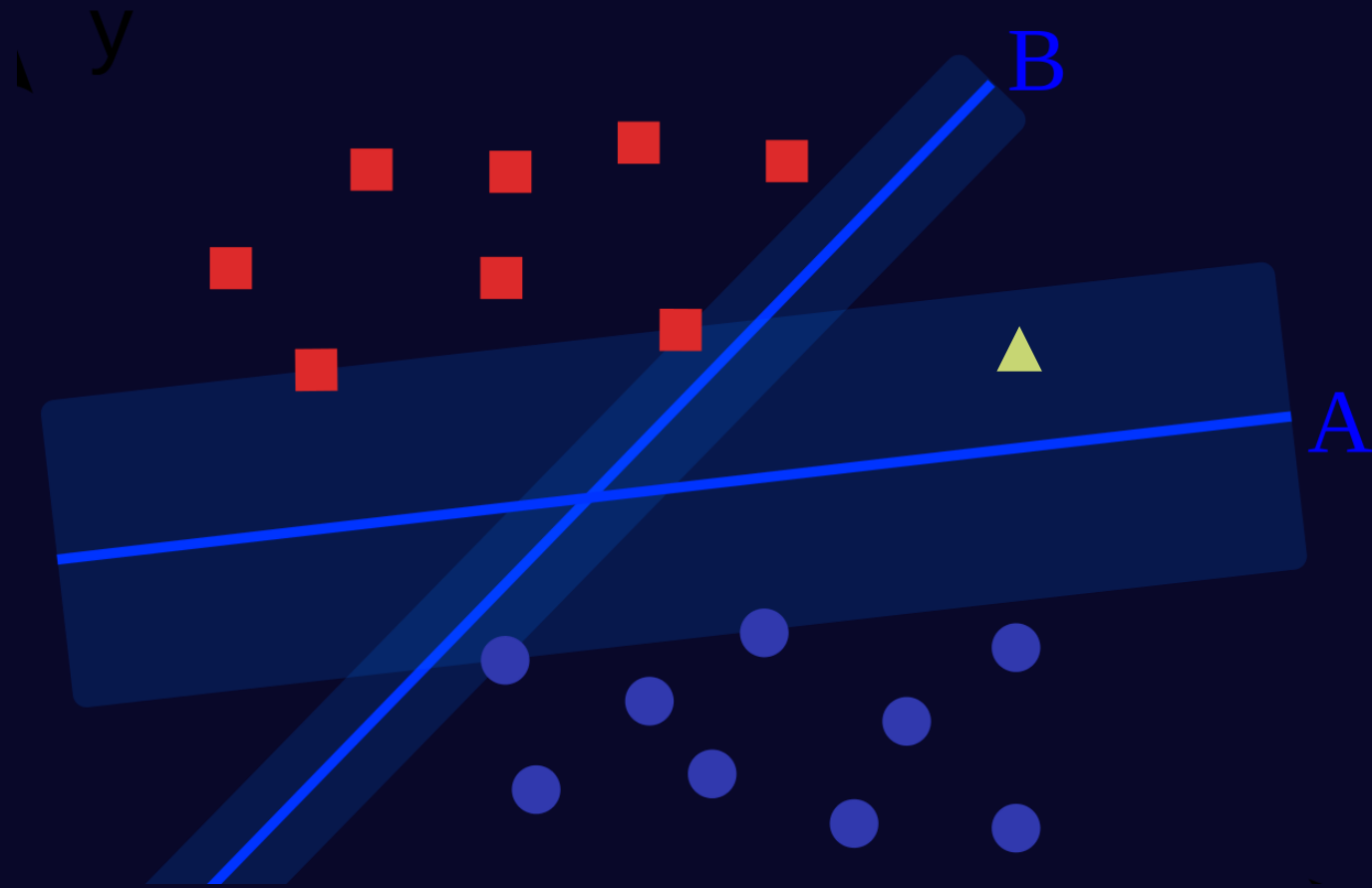


- Schnelle Verarbeitung, trotz riesiger Vektoren
- Kontextinformationen gehen verloren:

“I love Paris but I hate New York”

“I love New York but I hate Paris”

## Support Vector Machine: Bag-of-Word



# Support Vector Machine: Hyperebene

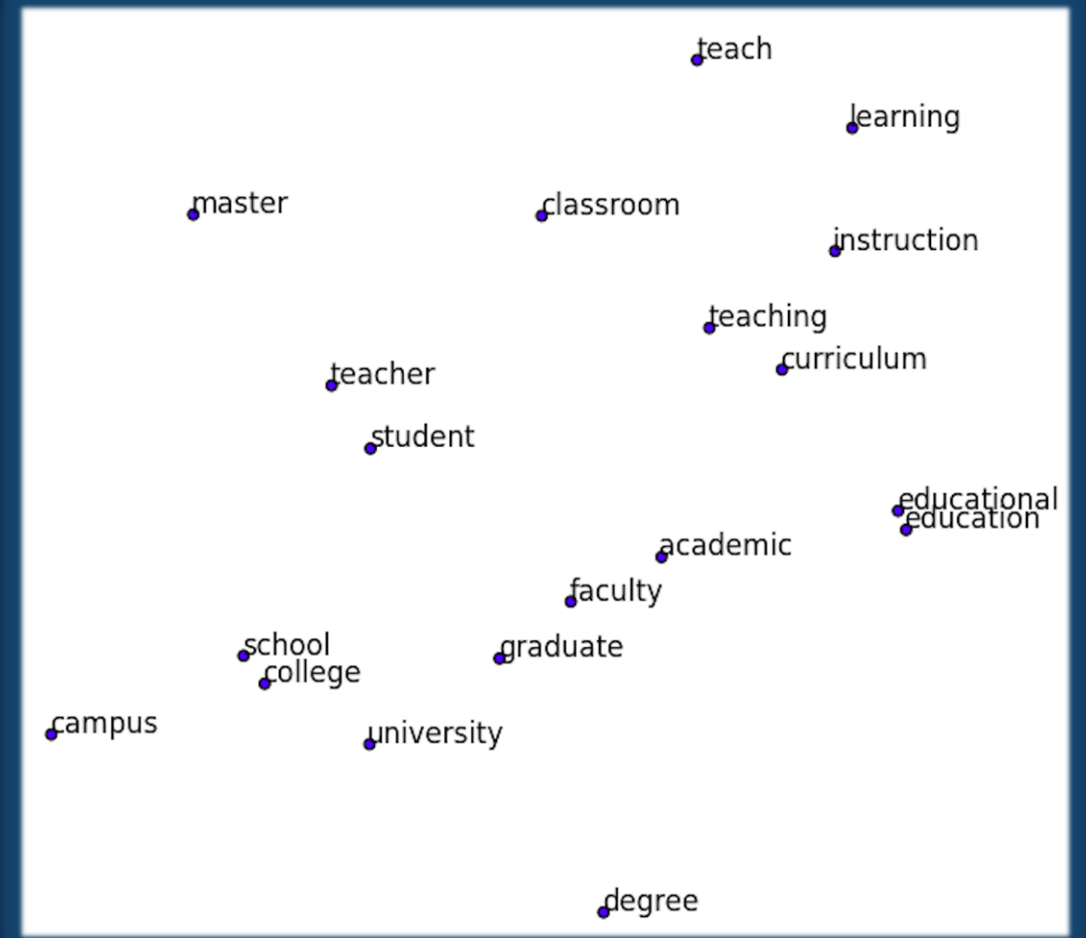
	1	2	3	4	5	6	7	8	9
man	1	0	0	0	0	0	0	0	0
woman	0	1	0	0	0	0	0	0	0
boy	0	0	1	0	0	0	0	0	0
girl	0	0	0	1	0	0	0	0	0
prince	0	0	0	0	1	0	0	0	0
princess	0	0	0	0	0	1	0	0	0
queen	0	0	0	0	0	0	1	0	0
king	0	0	0	0	0	0	0	1	0
monarch	0	0	0	0	0	0	0	0	1

	Femininity	Youth	Royalty
Man	0	0	0
Woman	1	0	0
Boy	0	1	0
Girl	1	1	0
Prince	0	1	1
Princess	1	1	1
Queen	1	0	1
King	0	0	1
Monarch	0.5	0.5	1

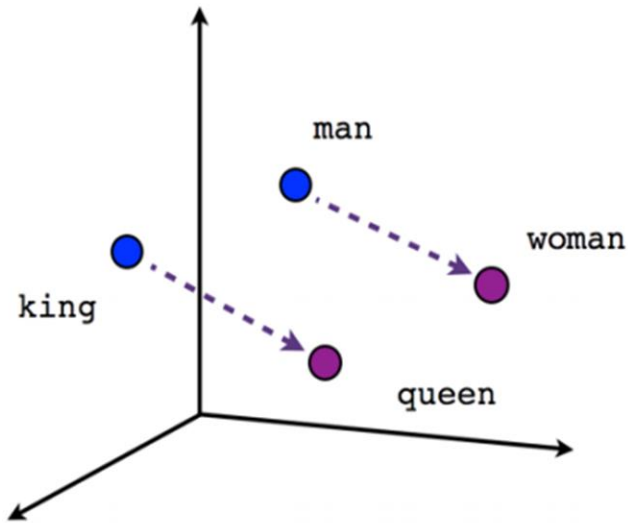
# Deep Learning: Word Embeddings

- Wörter, die im gleichen Kontext vorkommen, werden durch Vektoren in unmittelbarer Nähe zueinander dargestellt
- Training mit großen Korpora (Wikipedia)
- Jedes Wort erhält einen Zufallsvektor
- Vorhersage eines Wortes anhand Kontextwörter
- Anpassung der Vektorengewichte

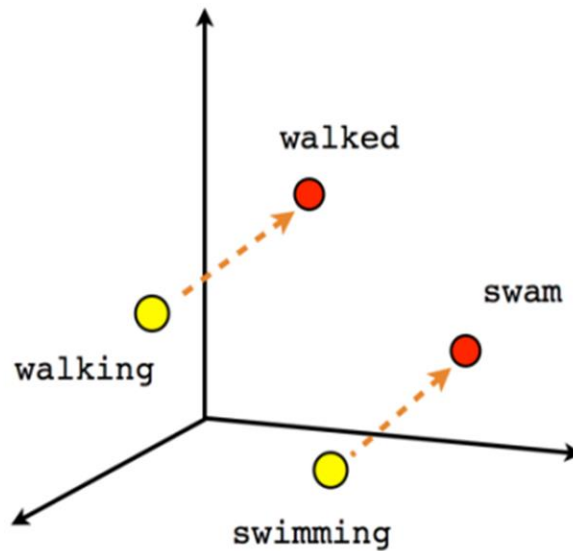
*“This is a patent document about lightbulbs”*



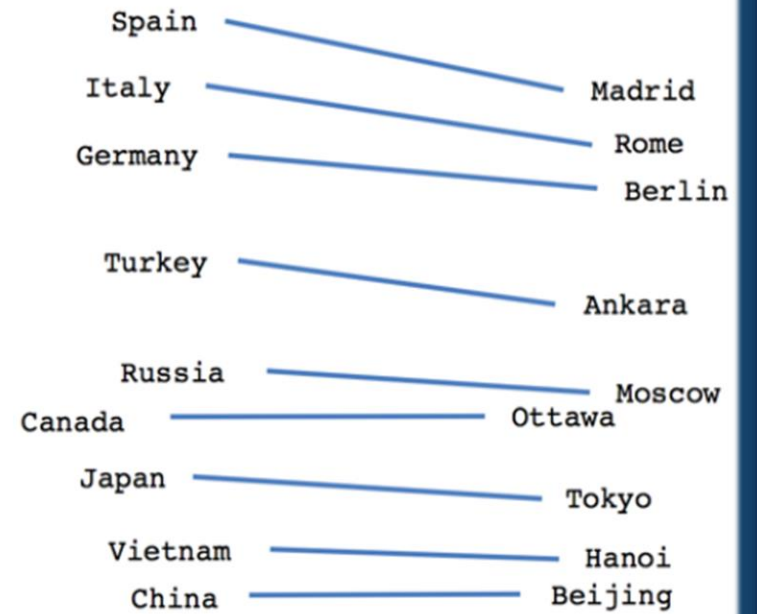
# Word Embeddings



Male-Female

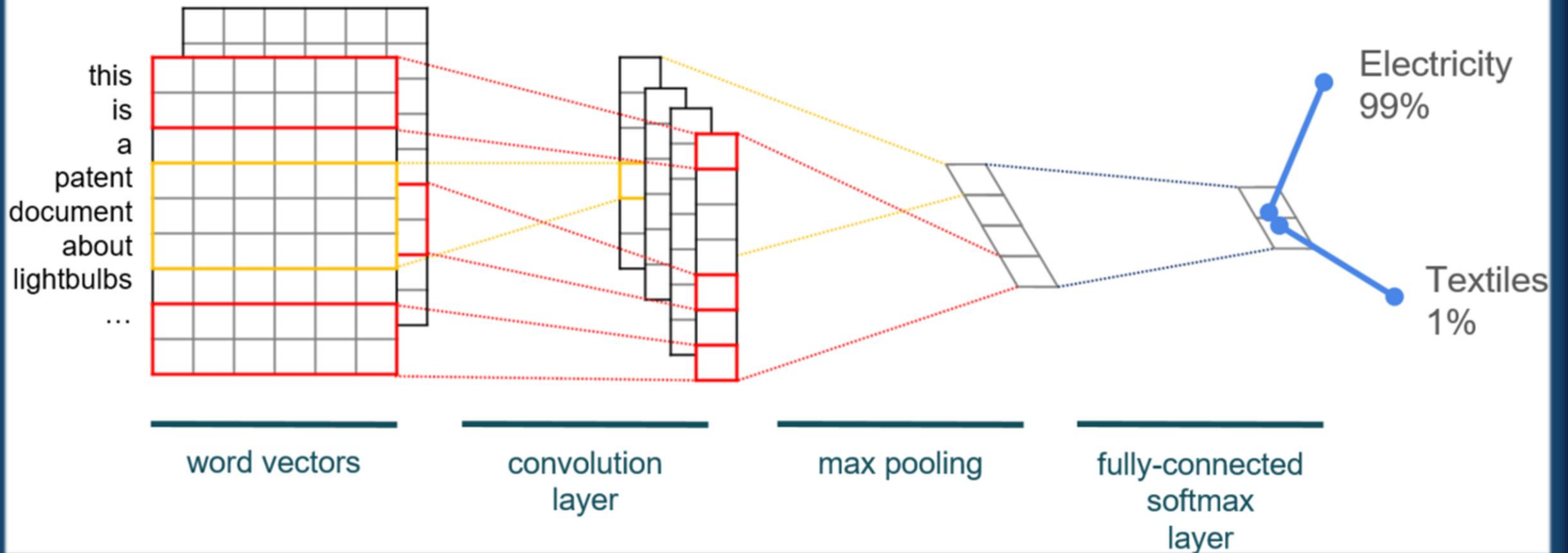


Verb tense



Country-Capital

# Eigenschaften von Word Embeddings



# Convolutional Neural Network (CNN)

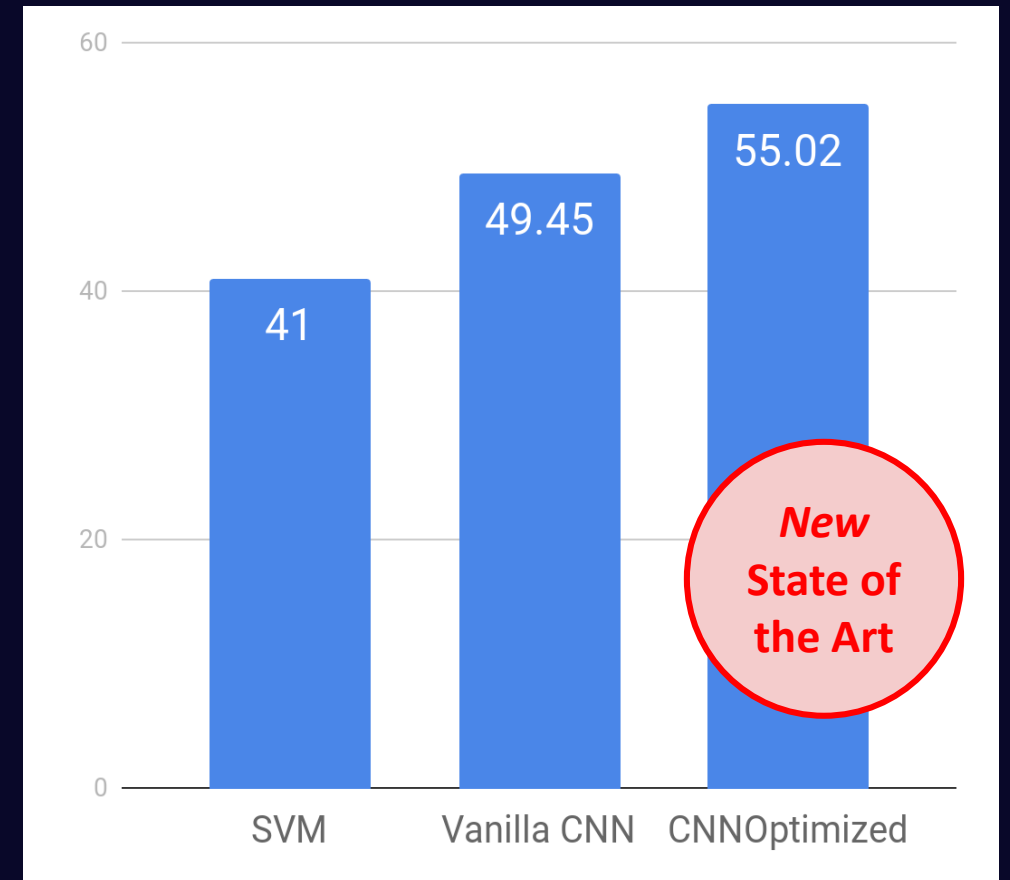
# Optimizing Neural Networks for Patent Classification

Louay Abdelgawad<sup>1</sup>✉, Peter Kluegl<sup>1</sup>, Erdan Genc<sup>1</sup>, Stefan Falkner<sup>2</sup>, and Frank Hutter<sup>2</sup>

<sup>1</sup> Averbis GmbH, Freiburg, Germany  
first.last@averbis.com

<sup>2</sup> Machine Learning Institute, Albert-Ludwigs University of Freiburg, Germany  
sfalkner@informatik.uni-freiburg.de  
fh@cs.uni-freiburg.de

- 75.000 Trainingsdokumente
- 23.000 Testdokumente
- 451 Zielkategorien



## Evaluation auf WIPO-Alpha Korpus

① Definition der Zielkategorien

② Bereitstellung von Beispielen / Trainieren

③ Klassifikation neuer Dokumente

GO

Feedback



**Aufsetzen einer neuen Anwendung**



Search and Classification 3d-Printing 1 Category: 3D-Bioprinting ✕ Q ✕

Category 594 results found Confidence (desc) Page size 5

Status Check All Label Documents Train Classify Accuracy: 86.5%

**Auto-Classified** 594

**Biocompatible crosslinked polymers** US-6566406-B1

**3D-Bioprinting ✕**

Assignee: INCEPT, LLC CPC Codes: C08G 63/912 12 Publication Date: 05/20/2003 Priority Date: 12/04/1998

**Abstract:**  
Biocompatible crosslinked polymers, and methods for their preparation and use, are disclosed in which the biocompatible crosslinked polymers are formed from water soluble precursors having electrophilic and nucleophilic groups capable of reacting and crosslinking in situ. Methods for making the resulting biocompatible crosslinked polymers biodegradable or not are provided, as are methods for controlling the rate of degradation. The crosslinking reactions may be carried out in situ on organs or tissues or outside the body. Applications for such biocompatible crosslinked polymers and their precursors include controlled delivery of drugs, prevention of post-operative adhesions, coating of medical devices such as vascular grafts, wound dressings and surgical sealants.

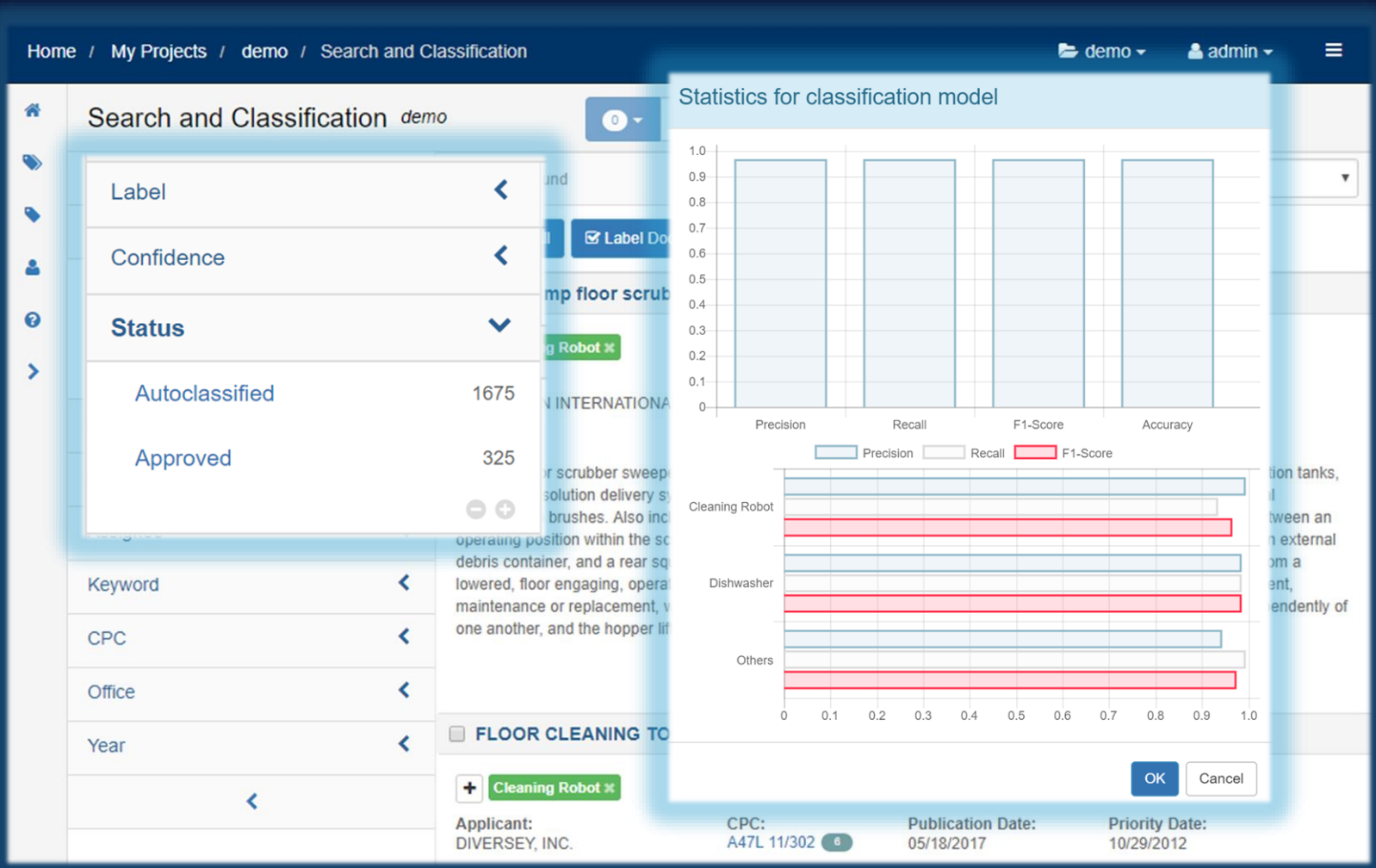
**CELL CULTURE** WO-2017068376-A1

**3D-Bioprinting ✕**

Assignee: - CPC Codes: B01L 2400/0475 8 Publication Date: 04/27/2017 Priority Date: 10/22/2015

**Abstract:**  
Aspects of the present invention relate to apparatus for use in cell and tissue culture techniques. Particularly, although not exclusively, embodiments of the present invention relate to apparatus which contribute to providing a dynamic cell culture environment. Also disclosed herein are methods for culturing cells and/or tissues, together with in vitro methods of testing drug efficacy as well as other subject matter.

# Maschinelles Lernen für Dokumentenklassifikation leicht gemacht



Automatische Klassifikation  
mit hoher Genauigkeit



**80% Zeitersparnis** bei der Identifizierung relevanter Dokumente



**Gleiche Genauigkeit** im Vergleich zur manuellen Klassifizierung durch IP-Experten



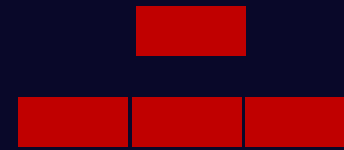
**bis zu 400% Verbesserung** gegenüber anderen automatisierten Diensten

**Vorteile maschinellen Lernens**

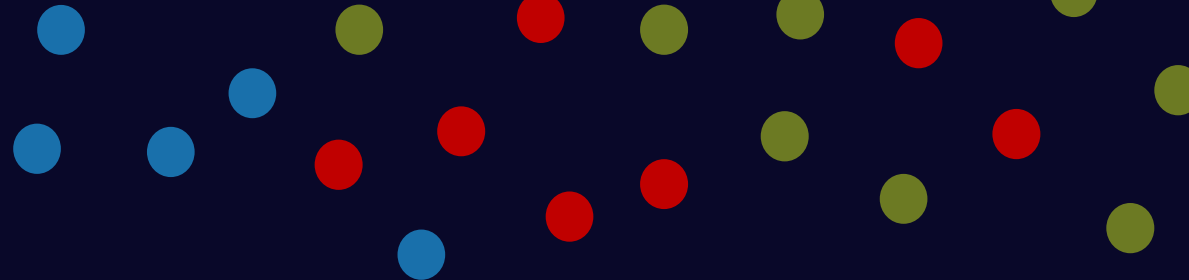
250 Abteilungen



~ 1.400 Bereiche



~ 250.000  
Patentanmeldungen



# Patentklassifikation @ EPO





Find. Understand. Predict.

Averbis GmbH  
Salzstraße 15  
79098 Freiburg



Kornél Markó  
[kornel.marko@averbis.com](mailto:kornel.marko@averbis.com)