



# Informationsextraktion aus Stellenanzeigen

Die Entwicklung maschineller Verfahren zur Extraktion von Kompetenzen

# Informationsextraktion – Grundlagen

## Aufgaben

- Auffinden relevanter Information in Texten
- Überführung der Information in strukturierte Repräsentation

## Ziele

- Nutzbare Organisation der Daten
- Ermöglichung der maschinellen Weiterverarbeitung

# Informationsextraktion – Teilbereiche

Named  
Entity  
Recognition

Markierung von Personen, Orten, Organisationen etc.

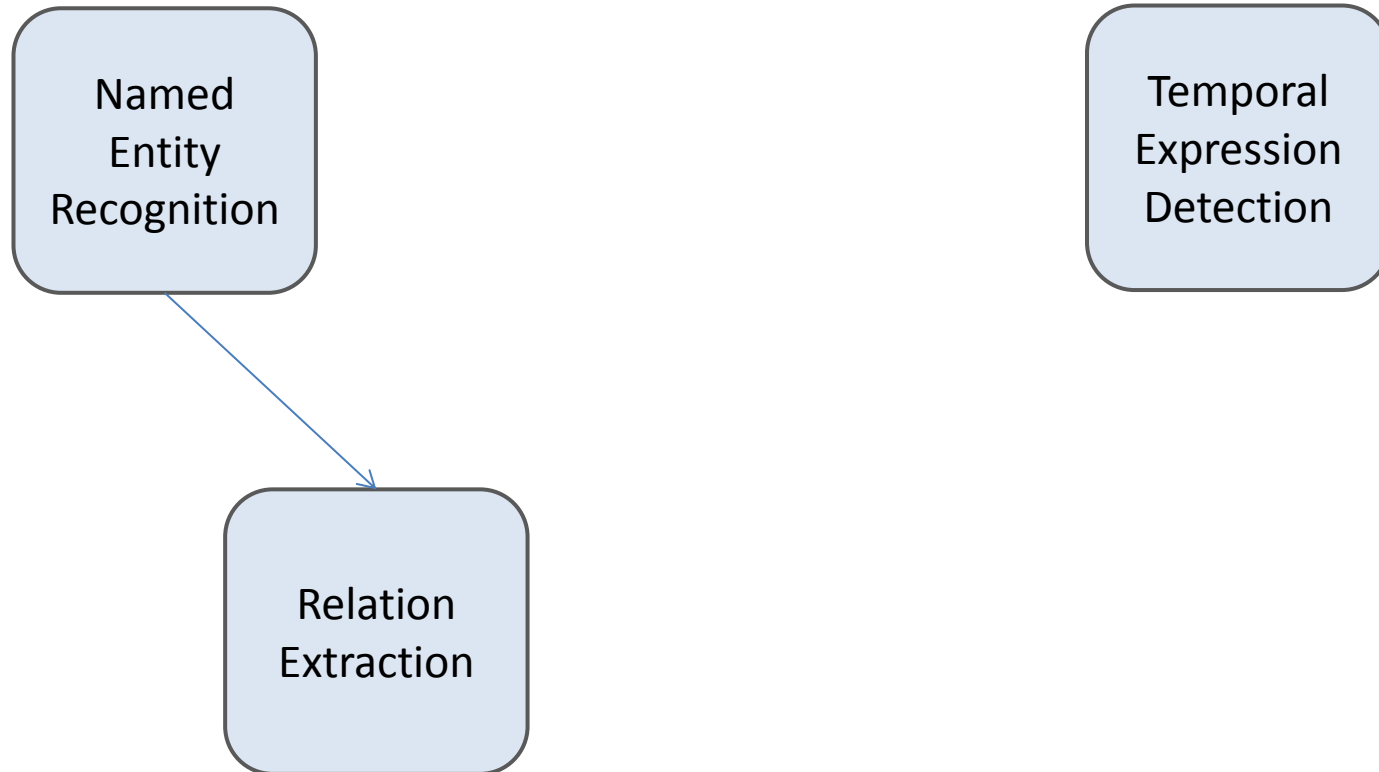
# Informationsextraktion – Teilbereiche

Named  
Entity  
Recognition

Temporal  
Expression  
Detection

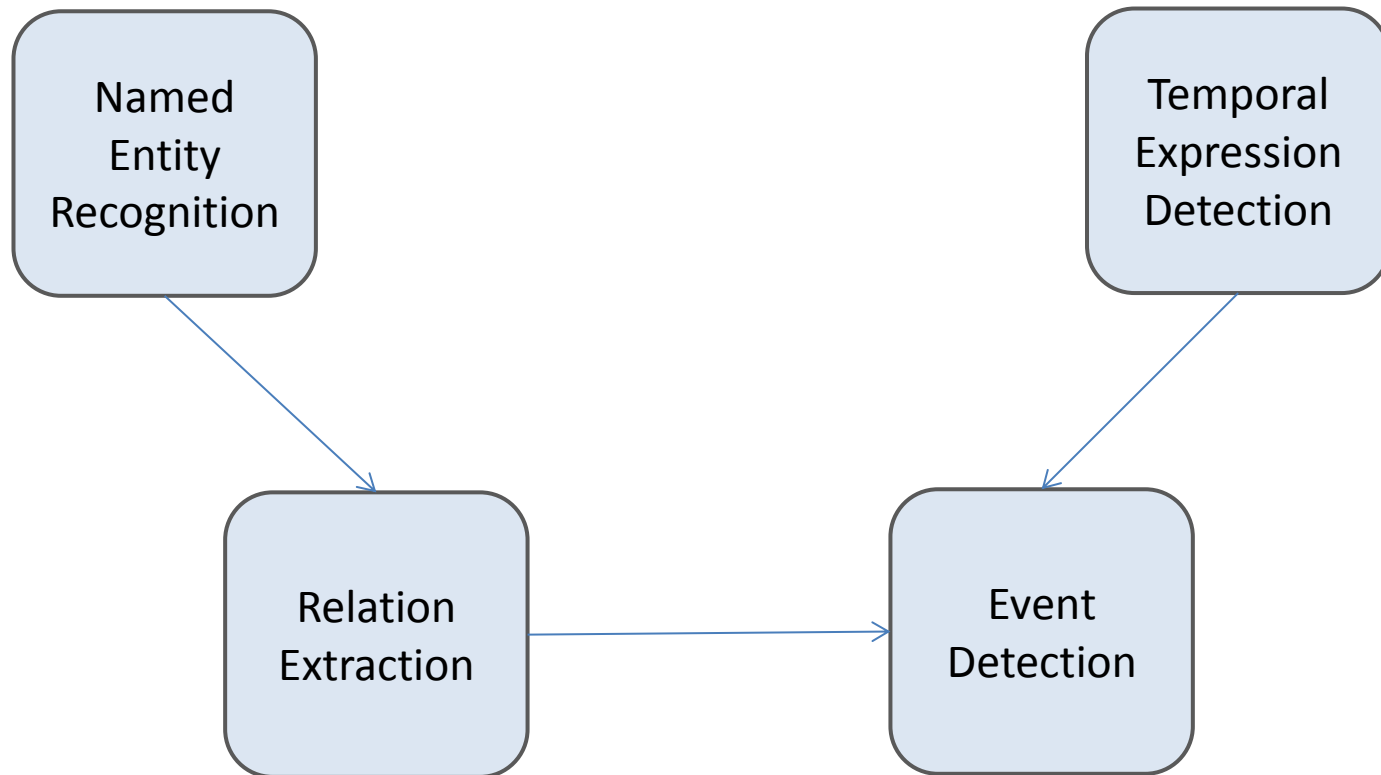
Markierung von absoluten (24. Dezember)  
und relativen (gestern) Zeitausdrücken

# Informationsextraktion – Teilbereiche



Interpretation von Relationen zwischen Named Entities,  
z.B. „X ist angestellt bei Y“ „Y hat Firmensitz in Z“

# Informationsextraktion – Teilbereiche



Interpretation von Ereignissen (z.B.  
„X hat Firma Y zum Termin Z verlassen“)

# Informationsextraktion – Stellenanzeigen

Arbeits-  
mittel

Welche Arbeitsmittel sollen im ausgeschriebenen Job eingesetzt werden?

# Informationsextraktion – Stellenanzeigen

Arbeits-  
mittel

Welche Arbeitsmittel sollen im ausgeschriebenen Job eingesetzt werden?

Tätigkeiten

Welche Tätigkeiten sollen im ausgeschriebenen Job ausgeführt werden?



# Informationsextraktion – Stellenanzeigen

Arbeits-  
mittel

Welche Arbeitsmittel sollen im ausgeschriebenen Job eingesetzt werden?

Tätigkeiten

Welche Tätigkeiten sollen im ausgeschriebenen Job ausgeführt werden?

Kompe-  
tenzen

Über welche Kompetenzen müssen / sollten Personen verfügen, die sich für den ausgeschriebenen Job bewerben?

# Informationsextraktion: Templates

Für die Leitung einer Filiale, eines der größten Sportfachhändler im Osten Deutschlands, suchen wir einen

**Teamleiter (m/w)**

Ihre Aufgabe

- Führung und Motivation der Mitarbeiter
- Ausbildung und Förderung der Auszubildenden und Mitarbeiter
- Personaleinsatzplanung
- Organisatin der Filialabläufe
- Verkauf von Sportartikeln

Sie

- überzeugen durch Kommunikationsfähigkeit
- denken und handeln eigenverantwortlich und unternehmerisch
- haben Sozialkompetenz und Teamgeist
- sind leistungsbereit und mobil

Bitte senden Sie ihre vollständigen Bewerbungsunterlagen ausschließlich per E-Mail #####@#####

Wert	Attribut
	Branche
	Beruf
	Ort
	Aufgaben
	Anforderungen

# Informationsextraktion: Templates

Für die Leitung einer Filiale, eines der größten Sportfachhändler im Osten Deutschlands, suchen wir einen

**Teamleiter (m/w)**

Ihre Aufgabe

- Führung und Motivation der Mitarbeiter
- Ausbildung und Förderung der Auszubildenden und Mitarbeiter
- Personaleinsatzplanung
- Organisatin der Filialabläufe
- Verkauf von Sportartikeln

Sie

- überzeugen durch Kommunikationsfähigkeit
- denken und handeln eigenverantwortlich und unternehmerisch
- haben Sozialkompetenz und Teamgeist
- sind leistungsbereit und mobil

Bitte senden Sie ihre vollständigen Bewerbungsunterlagen ausschließlich per E-Mail #####@#####

Wert	Attribut
	Branche
Teamleiter (m/w)	Beruf
	Ort
	Aufgaben
	Anforderungen

# Informationsextraktion: Templates

Für die Leitung einer Filiale, eines der größten Sportfachhändler im Osten Deutschlands, suchen wir einen

**Teamleiter (m/w)**

Ihre Aufgabe

- Führung und Motivation der Mitarbeiter
- Ausbildung und Förderung der Auszubildenden und Mitarbeiter
- Personaleinsatzplanung
- Organisatin der Filialabläufe
- Verkauf von Sportartikeln

Sie

- überzeugen durch Kommunikationsfähigkeit
- denken und handeln eigenverantwortlich und unternehmerisch
- haben Sozialkompetenz und Teamgeist
- sind leistungsbereit und mobil

Bitte senden Sie ihre vollständigen Bewerbungsunterlagen ausschließlich per E-Mail #####@#####

Wert	Attribut
	Branche
Teamleiter (m/w)	Beruf
	Ort
	Aufgaben
	Anforderungen

# Informationsextraktion: Templates

Für die Leitung einer Filiale, eines der größten Sportfachhändler im Osten Deutschlands, suchen wir einen

**Teamleiter (m/w)**

Ihre Aufgabe

- Führung und Motivation der Mitarbeiter
- Ausbildung und Förderung der Auszubildenden und Mitarbeiter
- Personaleinsatzplanung
- Organisatin der Filialabläufe
- Verkauf von Sportartikeln

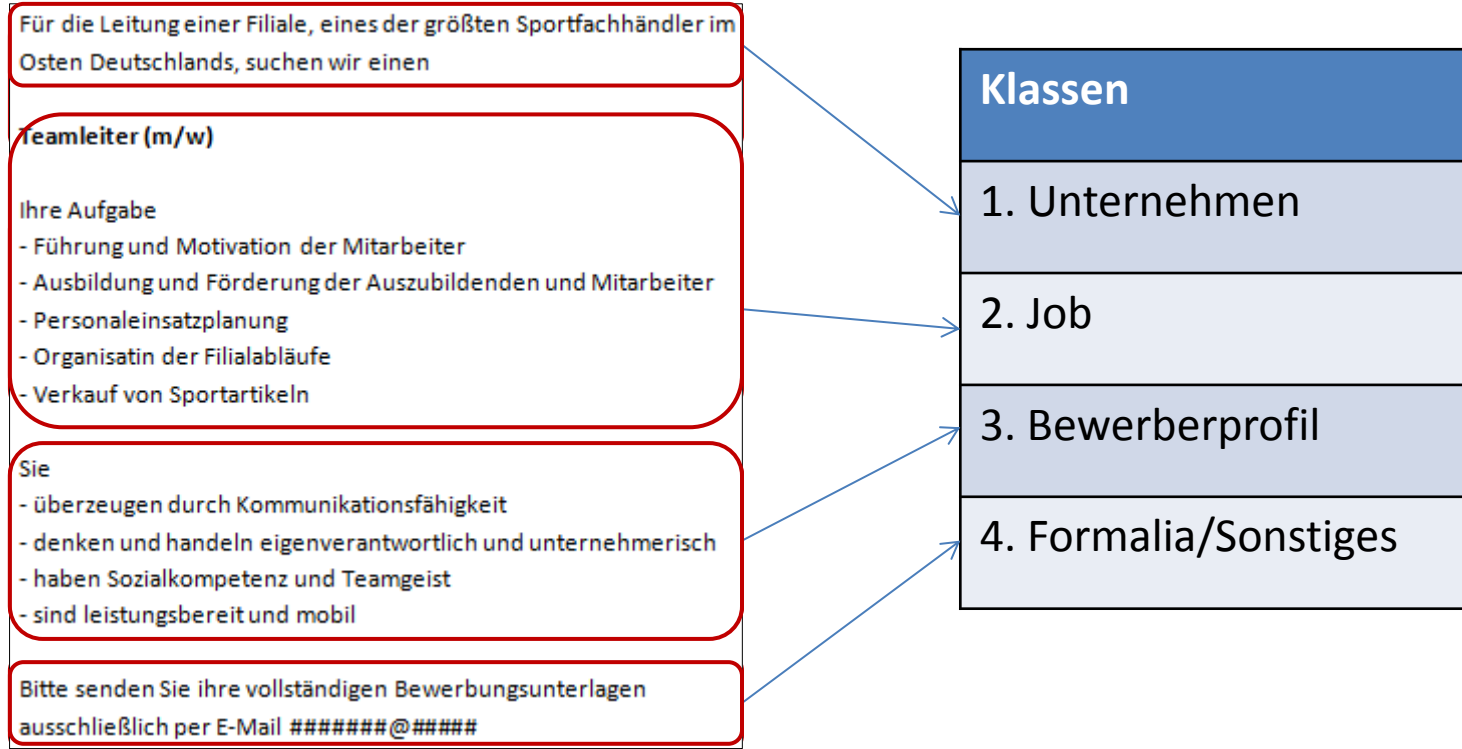
Sie

- überzeugen durch Kommunikationsfähigkeit
- denken und handeln eigenverantwortlich und unternehmerisch
- haben Sozialkompetenz und Teamgeist
- sind leistungsbereit und mobil

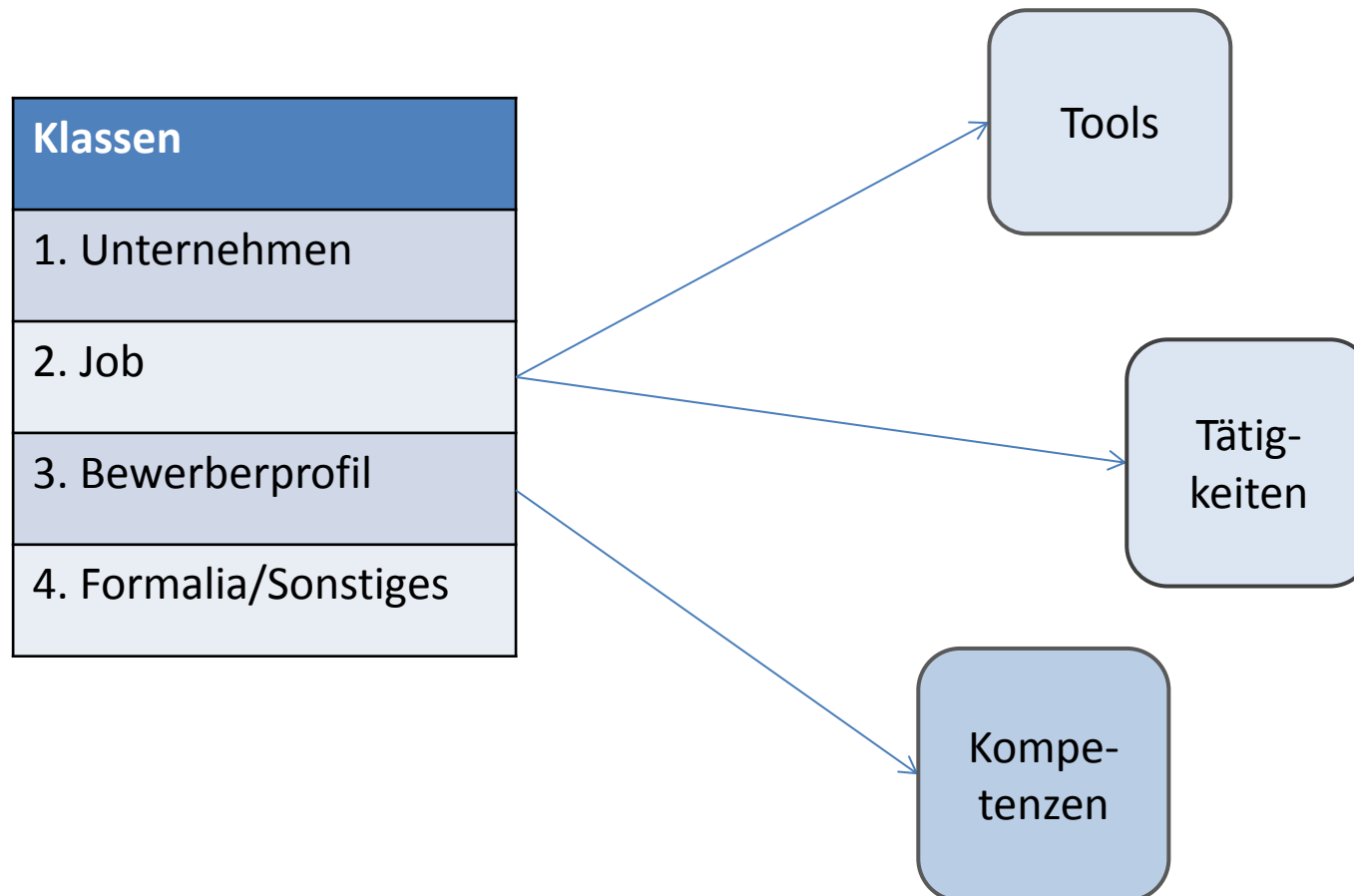
Bitte senden Sie ihre vollständigen Bewerbungsunterlagen ausschließlich per E-Mail #####@#####

Wert	Attribut
Sportfachhandel	Branche
Teamleiter (m/w)	Beruf
Ostdeutschland	Ort
Personaleinsatzplanung,...	Aufgaben
Teamgeist, Sozialkompetenz, ...	Anforderungen

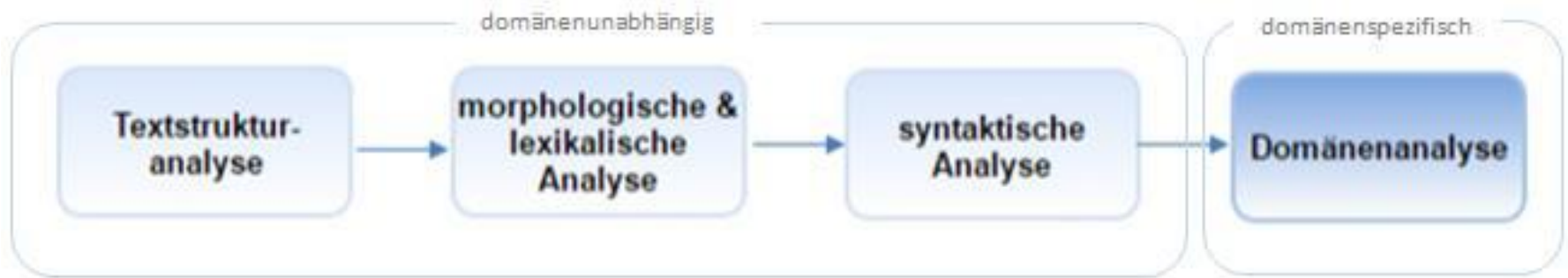
# Vorstufe: Klassifikation von Abschnitten



# Vorstufe: Klassifikation von Abschnitten

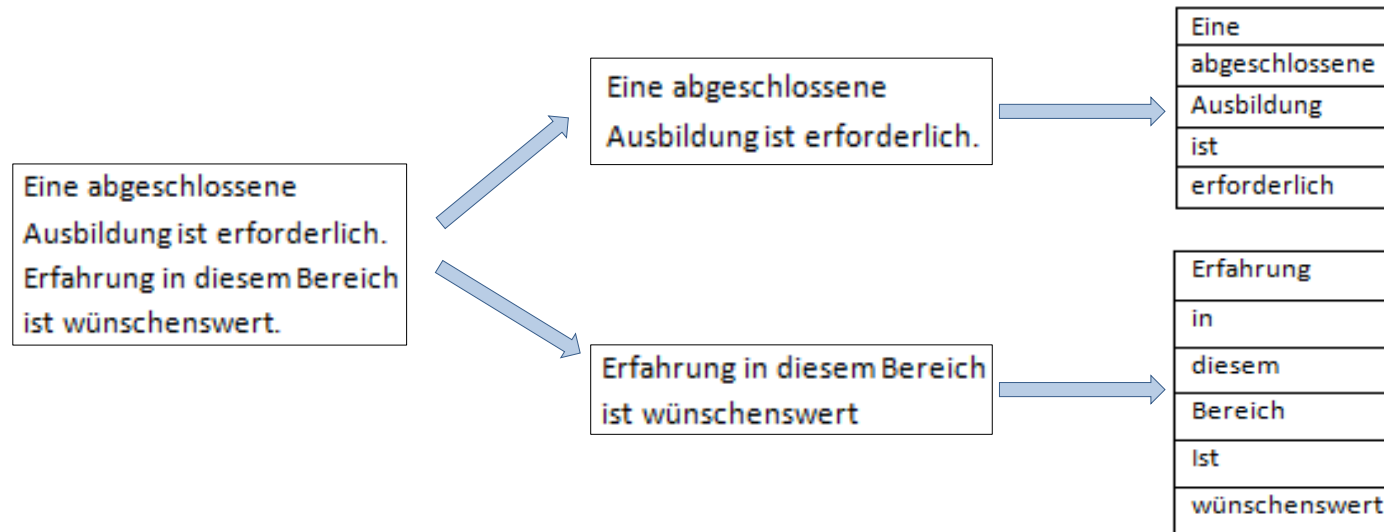
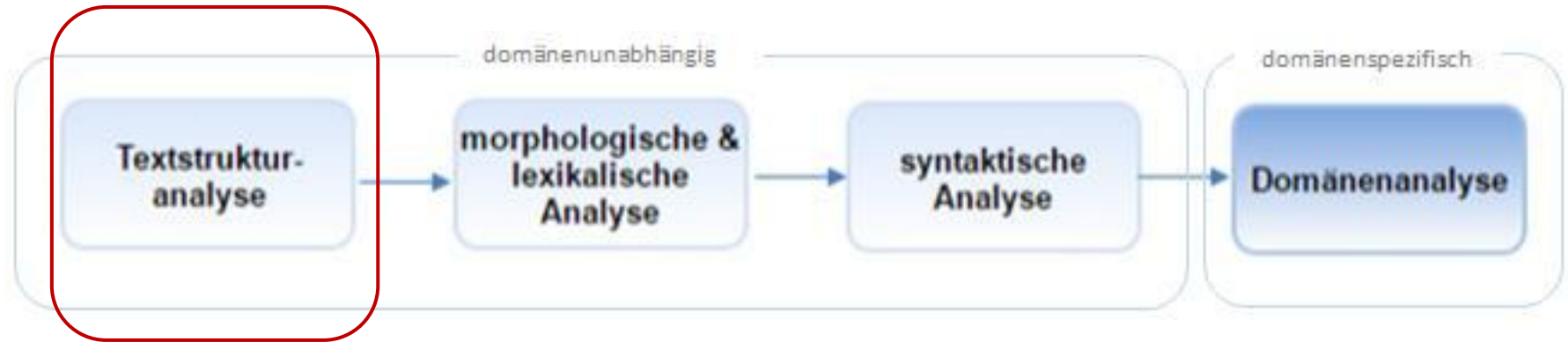


# Architektur eines IE-Systems

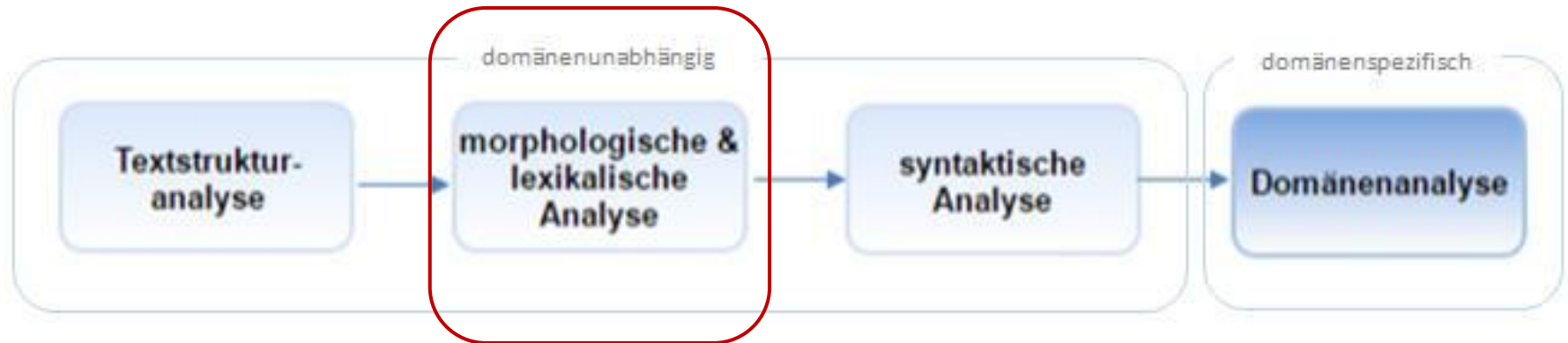




# Architektur eines IE-Systems



# Architektur eines IE-Systems

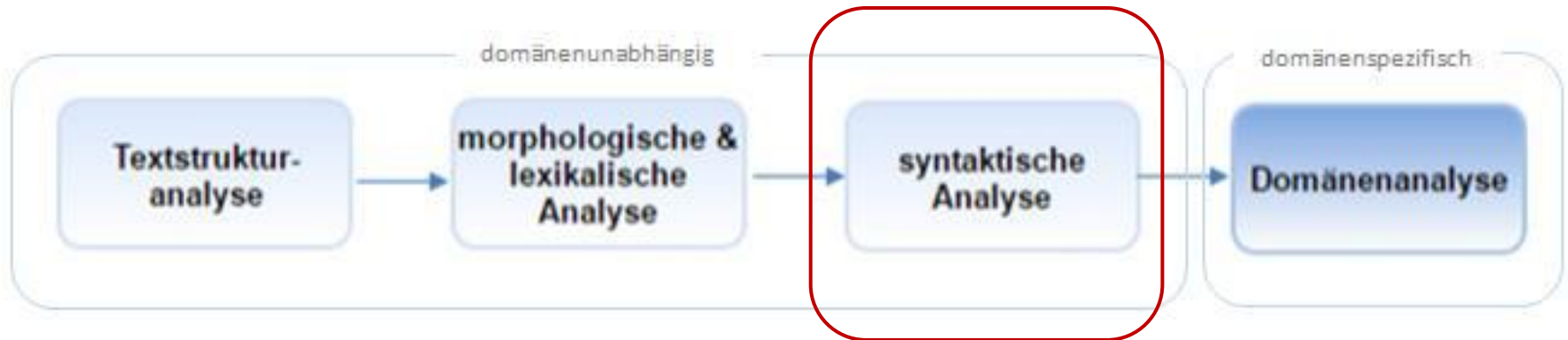


Erfahrung
in
diesem
Bereich
ist
wünschenswert

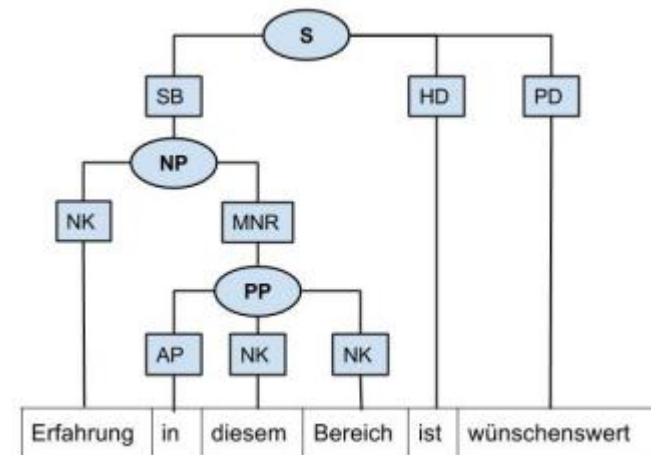


String	Lemma	POS
Erfahrung	erfahrung	N
in	in	P
diesem	dies	Art
Bereich	bereich	N
ist	sein	V
wünschenswert	wünschenswert	Adj

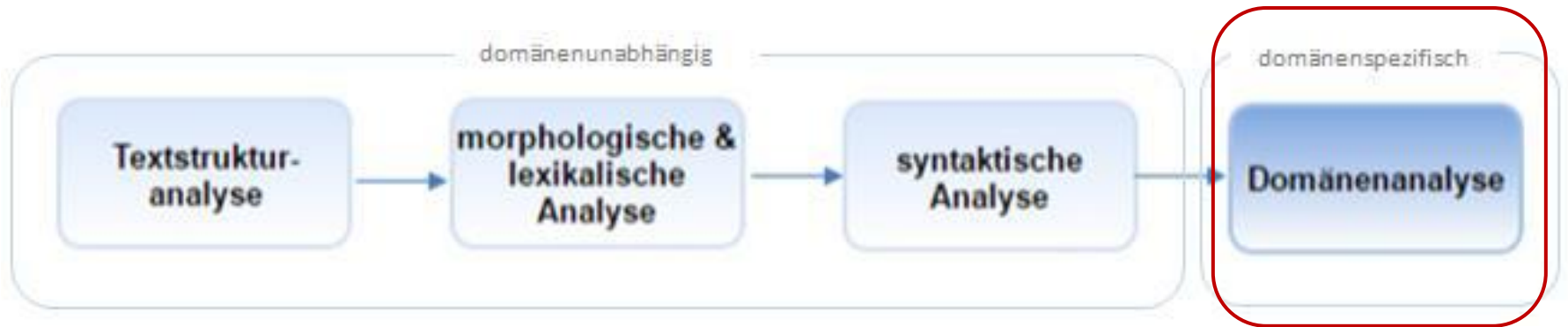
# Architektur eines IE-Systems



Erfahrung	erfahrung	N
in	in	P
diesem	dies	Art
Bereich	bereich	N
ist	sein	V
wünschenswert	wünschenswert	Adj

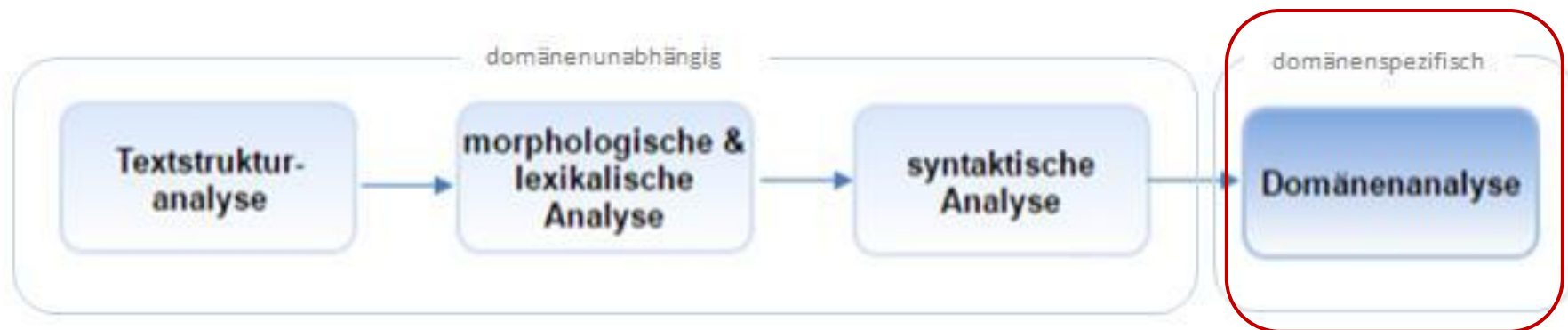


# Architektur eines IE-Systems



- Domänenspezifische Muster (Formulierungen) erkennen:
  - Führerschein **ist von Vorteil**
  - Spanischkenntnisse **wären von Vorteil**
  - Verkaufserfahrung **wäre von Vorteil**
  - **Kenntnisse in** Java
  - **Kenntnisse im** Programmieren
  - **Kenntnisse in** Photoshop


# Architektur eines IE-Systems




- Domänenspezifische Muster (Formulierungen) erkennen:

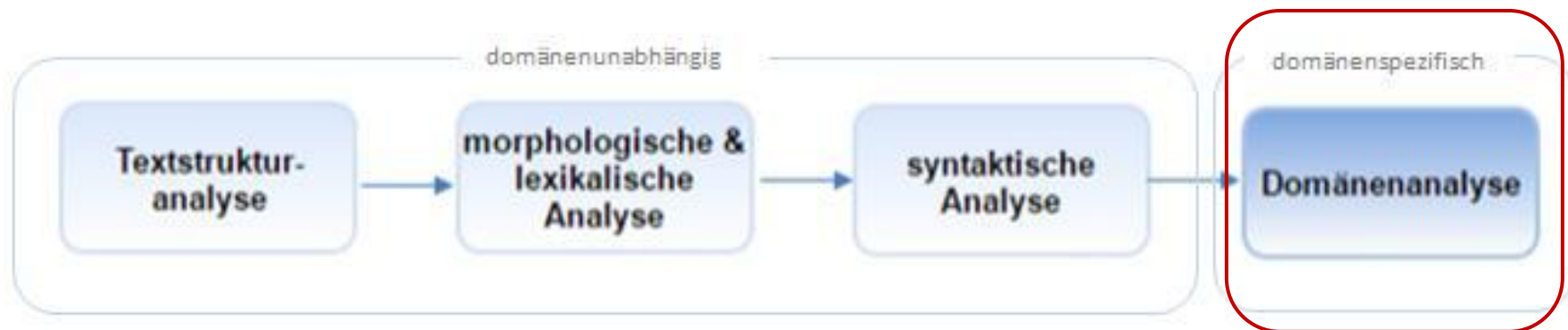
- Führerschein **ist von Vorteil**
- Spanischkenntnisse **wären von Vorteil**
- Verkaufserfahrung **wäre von Vorteil**
- Kenntnisse in Java**
- Kenntnisse im Programmieren**
- Kenntnisse in Photoshop**

- Extraktionsregeln bilden:

	sein	von	vorteil
N	-	-	-

kenntnis	in	
-	-	N

# Architektur eines IE-Systems




- Domänenspezifische Muster (Formulierungen) erkennen:

- Führerschein **ist von Vorteil**
- Spanischkenntnisse **wären von Vorteil**
- Verkaufserfahrung **wäre von Vorteil**

- **Kenntnisse in** Java
- **Kenntnisse im** Programmieren
- **Kenntnisse in** Photoshop


- Extraktionsregeln bilden:

	sein	von	vorteil
N	-	-	-

kenntnis	in	
-	-	N


- Extraktionsregeln anwenden

# Domänenanalyse

kenntnis	in	
-	-	N




Sie sollten über gute Kenntnisse in **ms-office** und excel verfügen  
Wenn Sie bereits Grundkenntnisse im **Programmieren** haben, dann...

	sein	von	vorteil
N	-	-	-




**Berufserfahrung** ist von Vorteil  
**Französischkenntnisse** sind von Vorteil

sie	sein	
-	-	Adj



Sie sind **flexibel**, belastbar und motiviert? Dann...

# Domänenanalyse

kenntnis	in	
-	-	N




Sie sollten über gute Kenntnisse in **ms-office** und excel verfügen  
Wenn Sie bereits Grundkenntnisse im **Programmieren** haben, dann...

	sein	MODIFIER	
N	-	-	-



**Berufserfahrung** ist von Vorteil  
**Französischkenntnisse** sind von Vorteil


sie	sein	
-	-	Adj



Sie sind **flexibel**, belastbar und motiviert? Dann...



# Domänenanalyse

kenntnis	in	
-	-	N




Sie sollten über gute Kenntnisse in **ms-office** und excel verfügen  
Wenn Sie bereits Grundkenntnisse im **Programmieren** haben, dann...

	sein	MODIFIER	
N	-	-	-



**Berufserfahrung** ist von Vorteil  
**Französischkenntnisse** sind von Vorteil


**adr-schein** ist erforderlich  
Ein **Führerschein** wäre vorteilhaft

sie	sein	
-	-	Adj



Sie sind **flexibel**, belastbar und motiviert? Dann...

# Domänenanalyse

kenntnis	in	
-	-	N




Sie sollten über gute Kenntnisse in **ms-office** und excel verfügen  
Wenn Sie bereits Grundkenntnisse im **Programmieren** haben, dann...

	sein	MODIFIER	
N	-	-	-





**Berufserfahrung** ist von Vorteil  
**Französischkenntnisse** sind von Vorteil

**adr-schein** ist erforderlich  
Ein **Führerschein** wäre vorteilhaft


sie	sein	
-	-	Adj



Sie sind **flexibel**, belastbar und motiviert? Dann...

KOMPETENZ	,		-	
-	§,	Adj	Konj	Adj

# Domänenanalyse

kenntnis	in	
-	-	N




Sie sollten über gute Kenntnisse in **ms-office** und excel verfügen  
Wenn Sie bereits Grundkenntnisse im **Programmieren** haben, dann...

	sein	MODIFIER	
N	-	-	-





**Berufserfahrung** ist von Vorteil  
**Französischkenntnisse** sind von Vorteil

**adr-schein** ist erforderlich  
Ein **Führerschein** wäre vorteilhaft

sie	sein	
-	-	Adj



Sie sind **flexibel**, belastbar und motiviert? Dann...

KOMPETENZ	,		-	
-	§,	Adj	Konj	Adj



Sie sind **flexibel**, **belastbar** und **motiviert**? Dann...

# Zwei Ansätze zur Umsetzung eines IE-Systems:

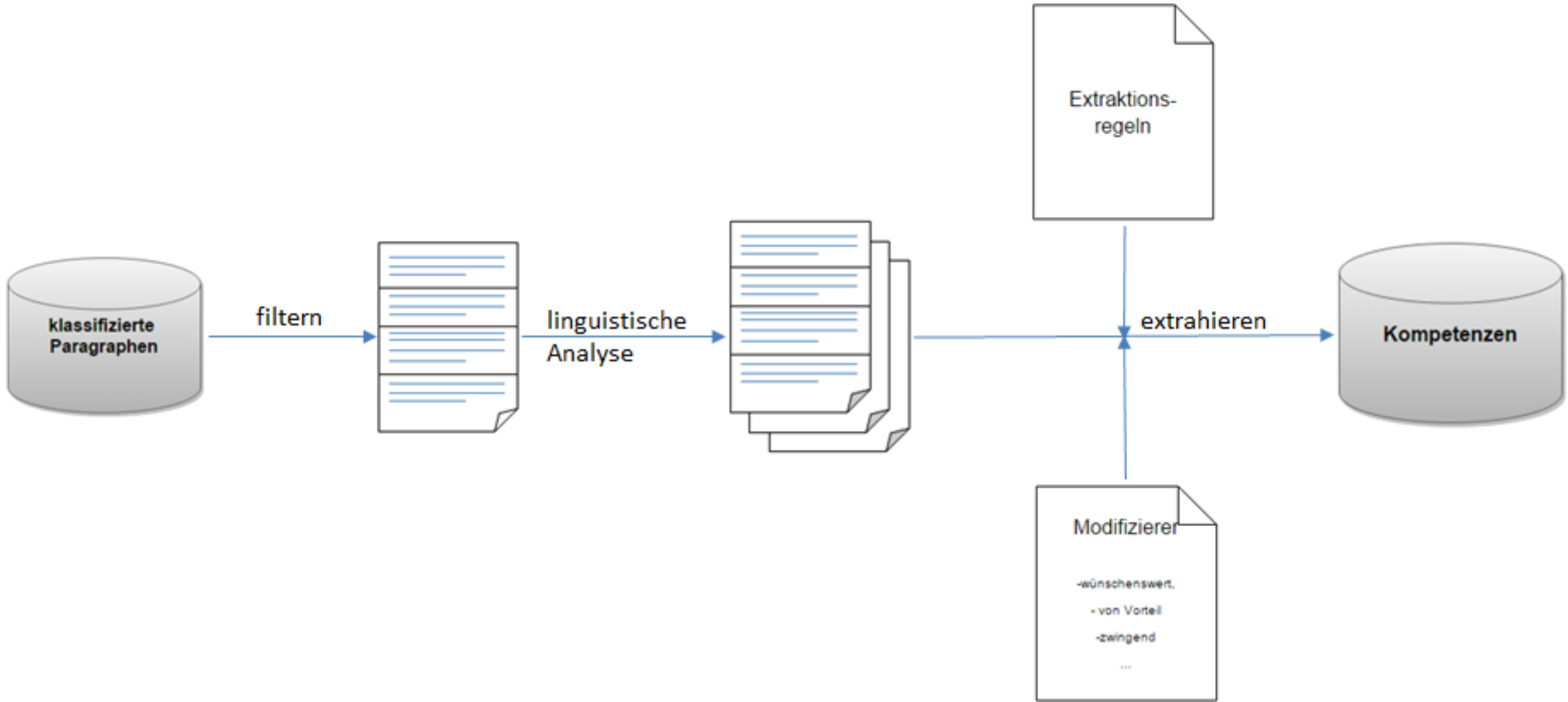
## 1. Knowledge Engineering

- Manuelle Formulierung von Extraktionsregeln
- Basierend auf beobachtbaren Mustern

## 2. Machine Learning

- Automatische Erkennung/Generierung von Extraktionsmustern
- Lernen aus Beispielen
- z.B. *Bootstrapping*

# Knowledge Engineering: Workflow



# Knowledge Engineering: Ergebnis

## Competences.db

Sentence	Comp	Importance	Contexts
Filter	Filter	Filter	Fi...
Wir erwarten Flexibilität, Teamgeist, Mobilität, gute PC-Kenntnisse, Interesse an Weiterbildung und moderner Technik.	gut pc-kenntnis	NULL	2
Gute Kommunikationsfähigkeiten und Serviceorientierung	gut kommunikationsfähigkeit	NULL	1
Gepflegtes Äußeres und Kommunikationsfähigkeit sind sehr wichtig	kommunikationsfähigkeit	NULL	1
Staplerschein wünschenswert	staplerschein	wünschenswert	1
Röntgenschein ist zwingend erforderlich.	röntgenschein	zwingend erforderlich	1
Zudem verfügen Sie über erste Berufserfahrung und die Bereitschaft sich in eine neue Aufgabe einzuarbeiten.	erster berufserfahrung	NULL	2
Deutsche Sprachkenntnisse, gute Umgangsformen und unbedingte Kundenorientierung sind für uns selbstverständlich.	deutsch sprachkenntnis	NULL	2
Studium der Fachrichtung Elektrotechnik oder ähnlicher Fachrichtung	studium der fachrichtung elek...	NULL	1
Alternativ: Ausbildung zum Sozialversicherungsfachangestellten (m_w), verbunden mit starker Affinität zur Informati...	ausbildung zu sozialversicher...	NULL	1
Sie haben eine abgeschlossene Ausbildung und bereits erste Berufserfahrung gesammelt.	abgeschlossen ausbildung	NULL	1
Sie haben eine abgeschlossene Ausbildung und bereits erste Berufserfahrung gesammelt.	erster berufserfahrung	NULL	1
Sie stehen kurz vor (oder nach) dem Abschluss Ihres Studiums der Informatik, Elektrotechnik oder eines vergleichba...	studium der informatik	NULL	1
Sie sind service- und kundenorientiert und haben eine gute Kommunikations - und Teamfähigkeit.	teamfähigkeit	NULL	1
Eine gute Chance haben Sie auch mit einem mittleren Bildungsabschluss mit anschließend abgeschlossener Berufsa...	abgeschlossen berufsausbildu...	NULL	1
Eine gute Chance haben Sie auch mit einem mittleren Bildungsabschluss mit anschließend abgeschlossener Berufsa...	mittlerer bildungsabschluss	NULL	1

# Knowledge Engineering: Evaluation

## Testkorpus

- ~ 550 Sätze
- ~ 450 manuell annotierte Kompetenzen
- ~ 40 Extraktionsmuster

## Ergebnis

- hohe Präzision
- geringe Ausbeute

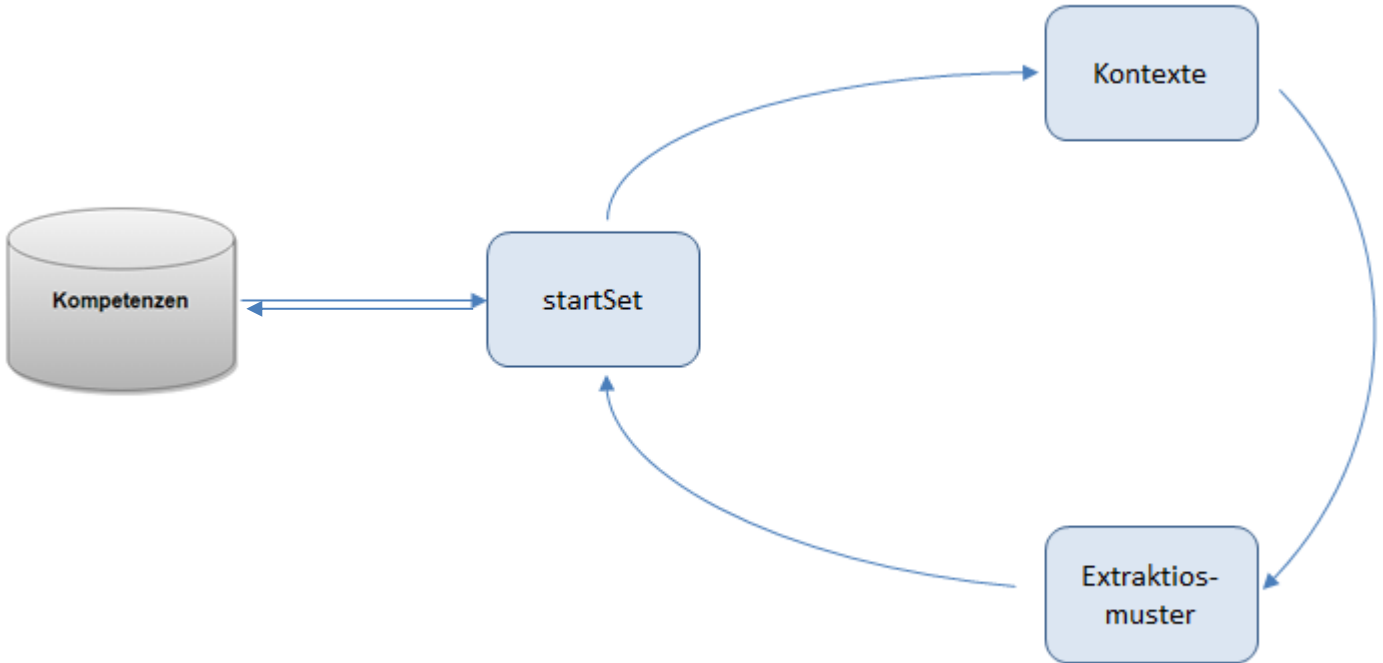
# Bootstrapping: Idee

*Nutze was du hat, um zu finden was du suchst!*

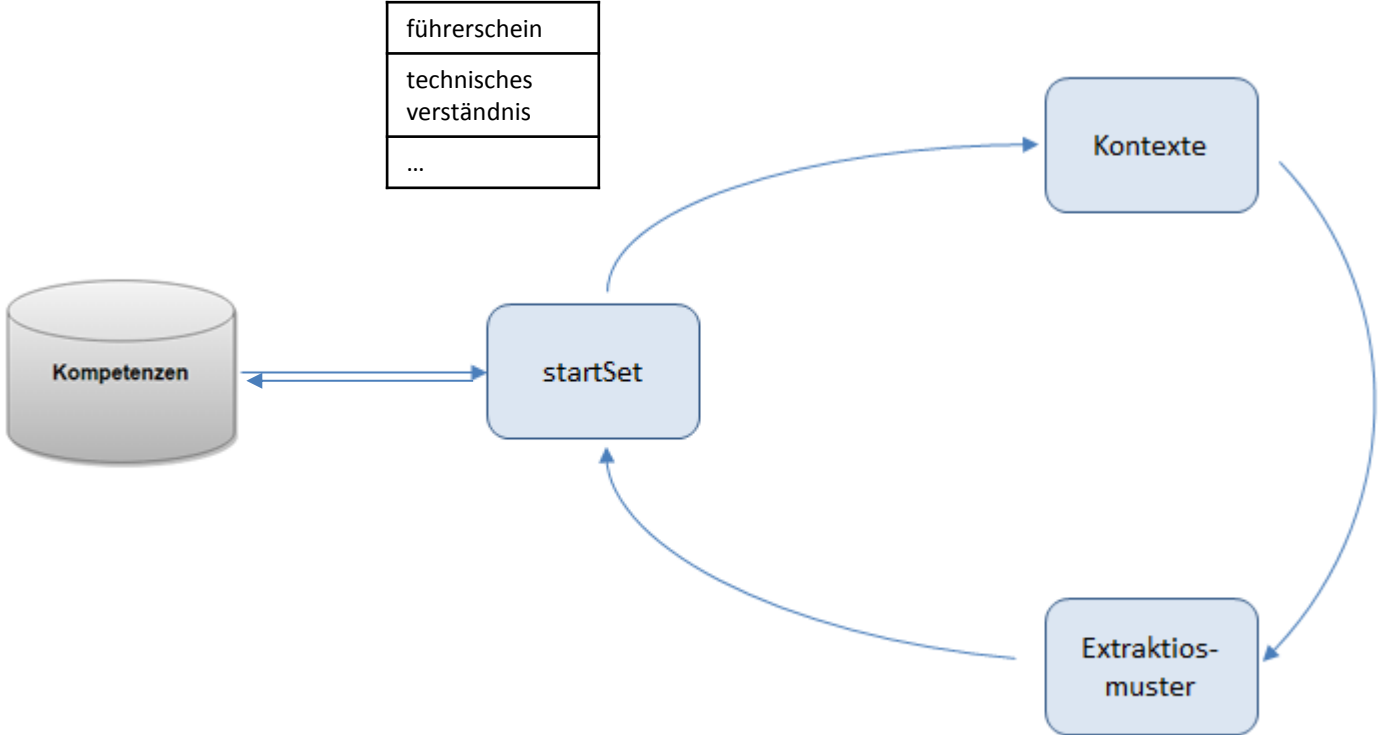
- Iteratives Verfahren:
  - Initiales Startset von Kompetenzausdrücken (Seed)
    - 1) Suche nach Vorkommen im Text
    - 2) Musterbildung durch Abstraktion der Kontexte
    - 3) Erweiterung des Startsets durch Anwendung der Kontexte



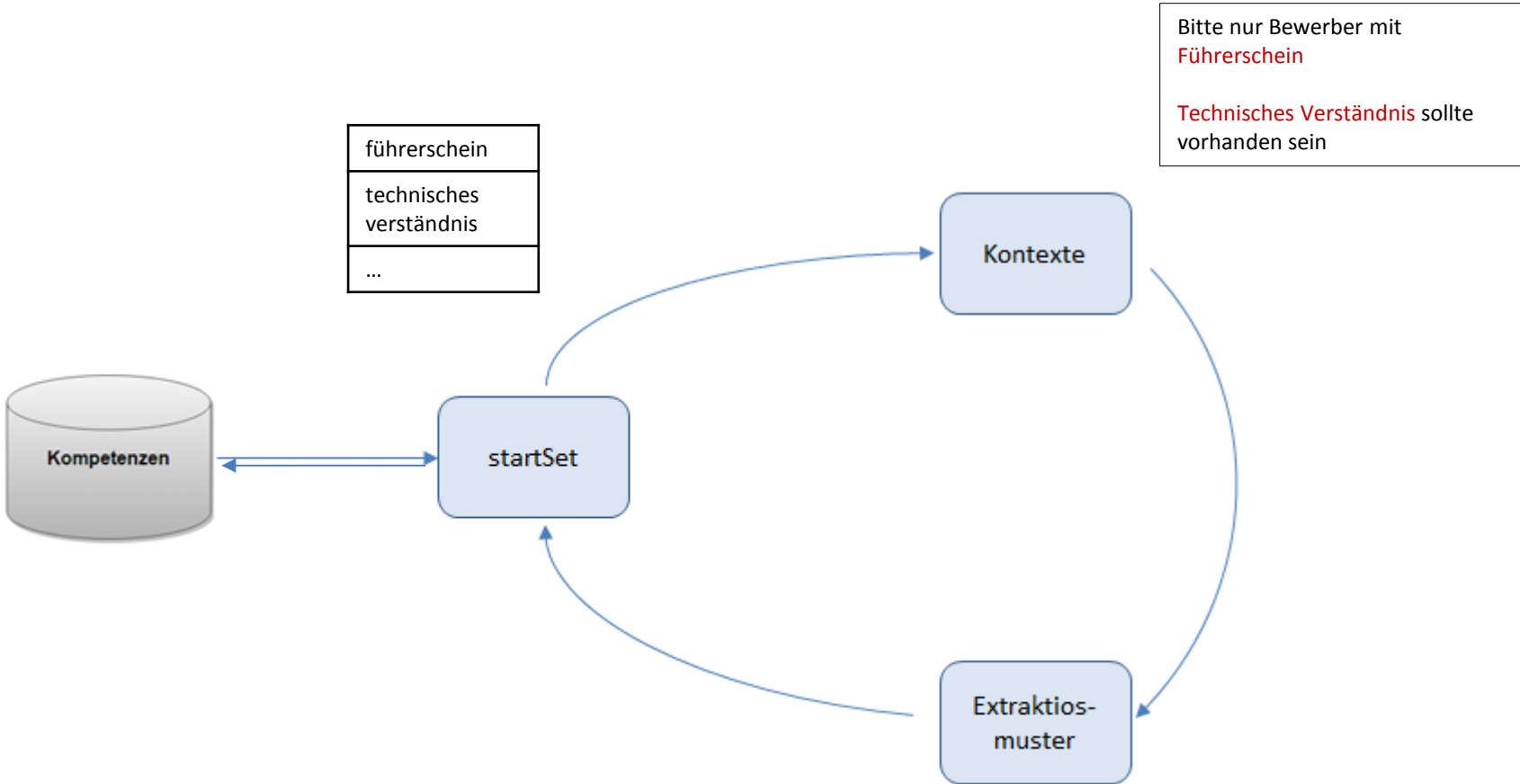
# Bootstrapping: Workflow



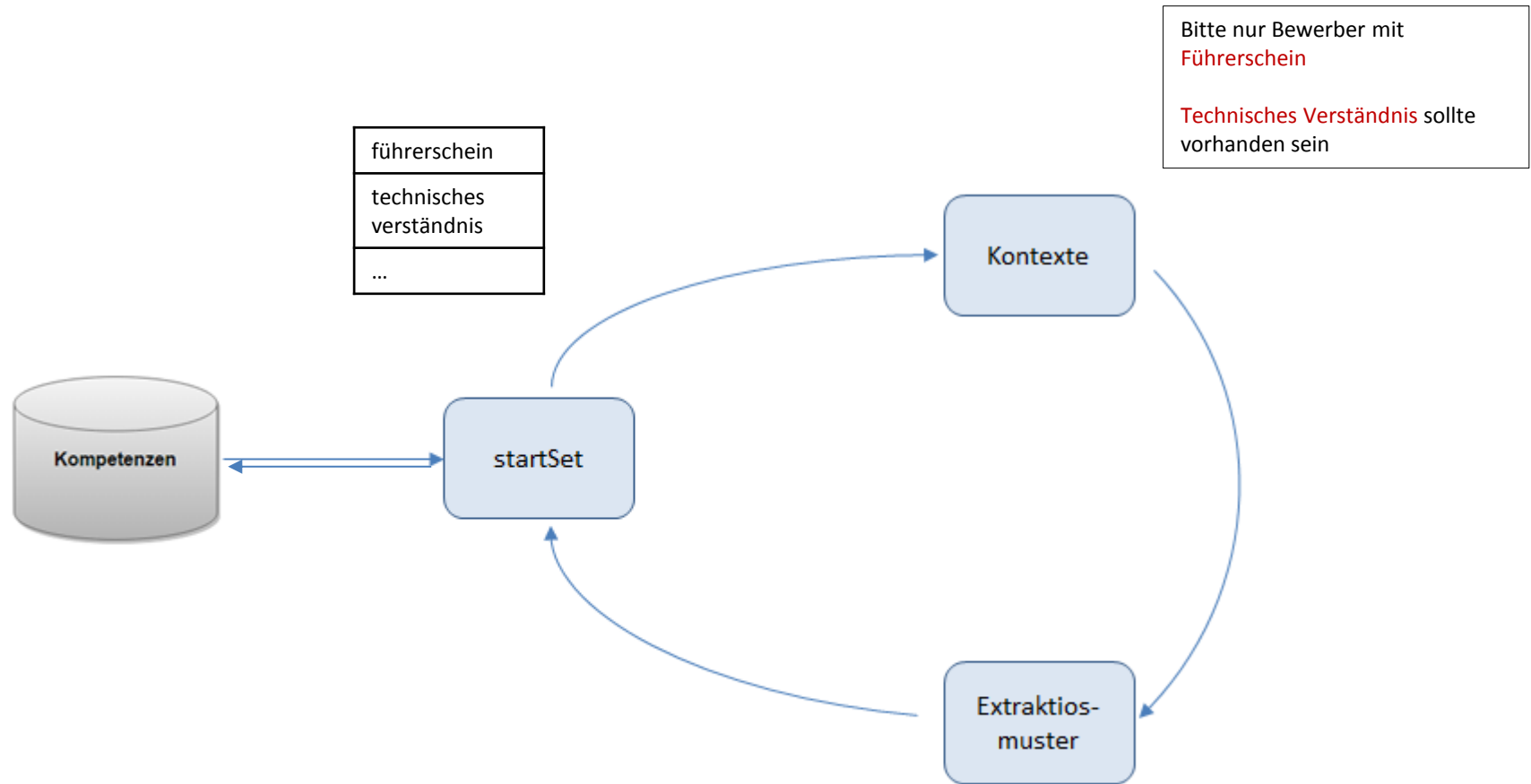
# Bootstrapping: Workflow



# Bootstrapping: Workflow



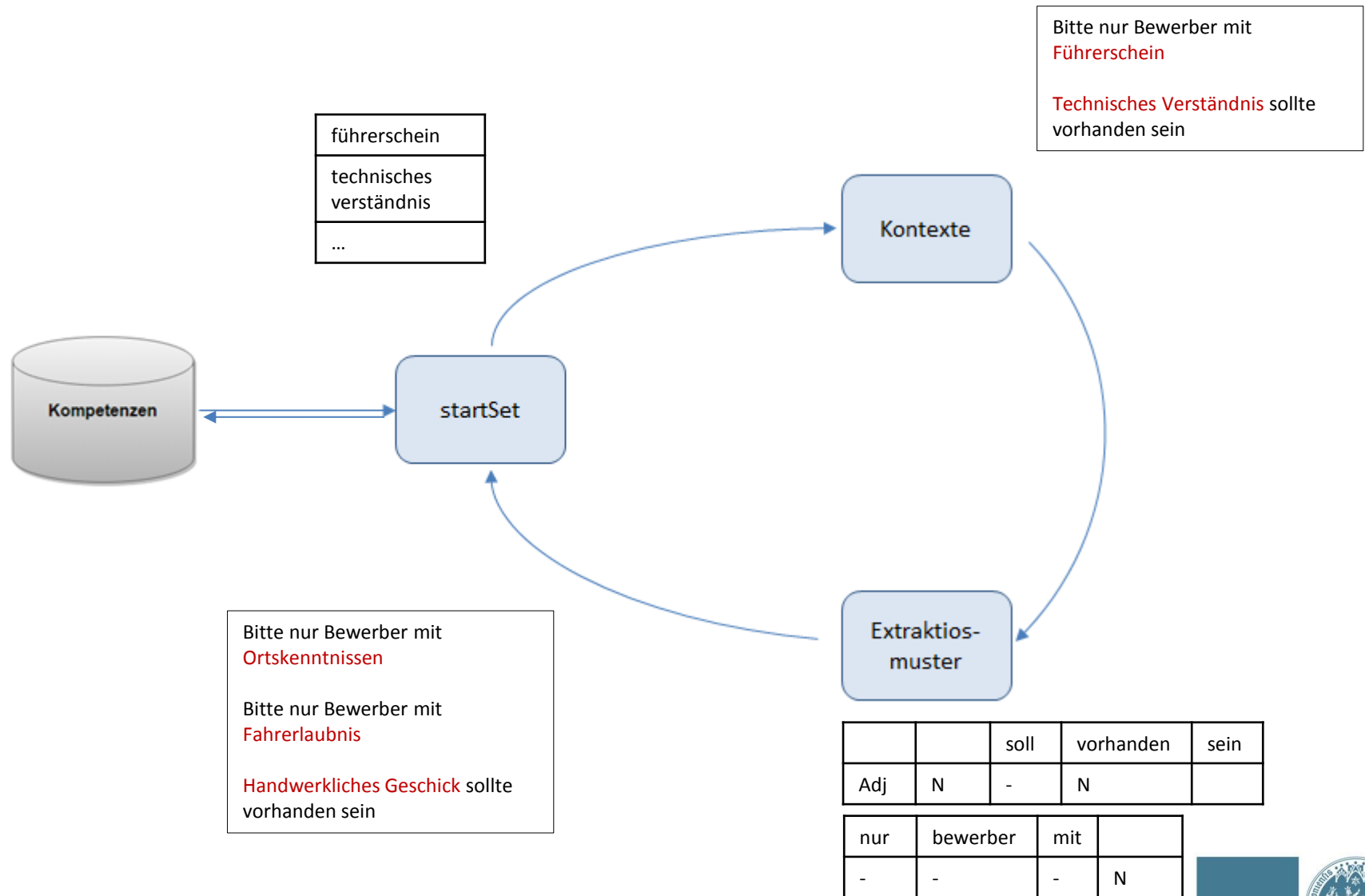
# Bootstrapping: Workflow



		soll	vorhanden	sein
Adj	N	-	N	

nur	bewerber	mit	
-	-	-	N

# Bootstrapping: Workflow



# Vielen Dank für die Aufmerksamkeit!



# Kompetenzen in Stellenanzeigen

Voraussetzung:

- Eigener Pkw, Führerschein und Kranschein **erforderlich**
- zuverlässig, flexibel, motiviert
- Erfahrung im Messebereich **wünschenswert aber nicht zwingend erforderlich**

Ihr Profil:

- Grundkenntnisse im Umgang mit ms-office
- Verantwortungsbewusstsein und handwerkliches Geschick
- Reisebereitschaft

Eine abgeschlossene Berufsausbildung ist **erforderlich**. Berufserfahrung wäre **von Vorteil**.

Sie sollten Kenntnisse in QuarkXpress, Photoshop und InDesign haben

# Kompetenzen in Stellenanzeigen

Voraussetzung:

- Eigener Pkw, Führerschein und Kranschein **erforderlich**
- zuverlässig, flexibel, motiviert
- **Erfahrung im** Messebereich **wünschenswert aber nicht zwingend erforderlich**

Ihr Profil:

- **Grundkenntnisse im Umgang mit** ms-office
- Verantwortungsbewusstsein und handwerkliches Geschick
- Reisebereitschaft

Eine abgeschlossene Berufsausbildung ist **erforderlich**. Berufserfahrung wäre **von Vorteil**.

Sie sollten **Kenntnisse in** QuarkXpress, Photoshop und InDesign haben



# Kompetenzen in Stellenanzeigen

## Voraussetzung:

- Eigener Pkw, Führerschein und Kranschein erforderlich
- zuverlässig, flexibel, motiviert
- Erfahrung im Messebereich wünschenswert aber nicht zwingend erforderlich

## Ihr Profil:

- Grundkenntnisse im Umgang mit ms-office
- Verantwortungsbewusstsein und handwerkliches Geschick
- Reisebereitschaft

Eine abgeschlossene Berufsausbildung ist erforderlich. Berufserfahrung wäre von Vorteil.

Sie sollten Kenntnisse in QuarkXpress, Photoshop und InDesign haben

# Kompetenzen in Stellenanzeigen

Sie haben mit Charme und Pfiff ihren Schreibtisch im Griff.  
Ihr Engagement kommt aus dem Herzen und Sie lassen sich auch von urlaubsreifen Mitmenschen nicht auf die Palme bringen.

Stets gut gelaunt behalten Sie auch in stressigen Situationen einen kühlen Kopf

Zudem besitzen Sie die Kompetenz, erklärungsbedürftige Sachverhalte zu analysieren und schriftlich gut strukturiert und verständlich aufzubereiten.

Darüber hinaus sind Sie in der Lage auch in schwierigen Situationen den Überblick und die Ruhe zu bewahren