UNIVERSITÄT BAYREUTH

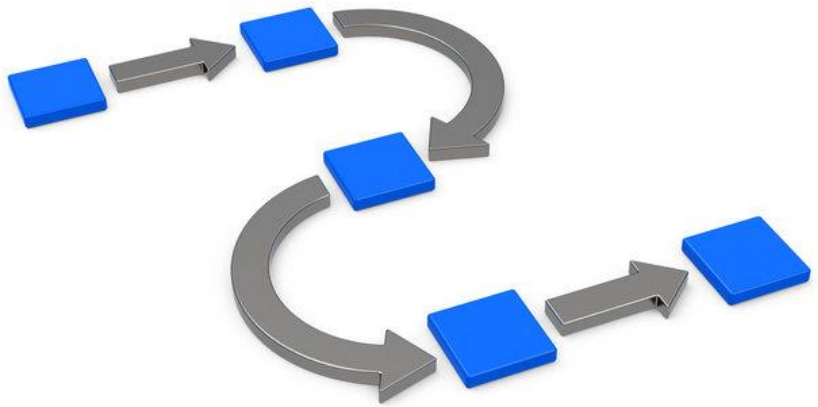
Automatisierte Datenextraktion auf Basis regulärer Ausdrücke

Am Beispiel von Werkzeugdaten

Johannes Mohr, Andreas Kormann,
Peter Grohmann, Stephan Tremmel

24. Bayreuther 3D-Konstrukteurstag

13. September 2023



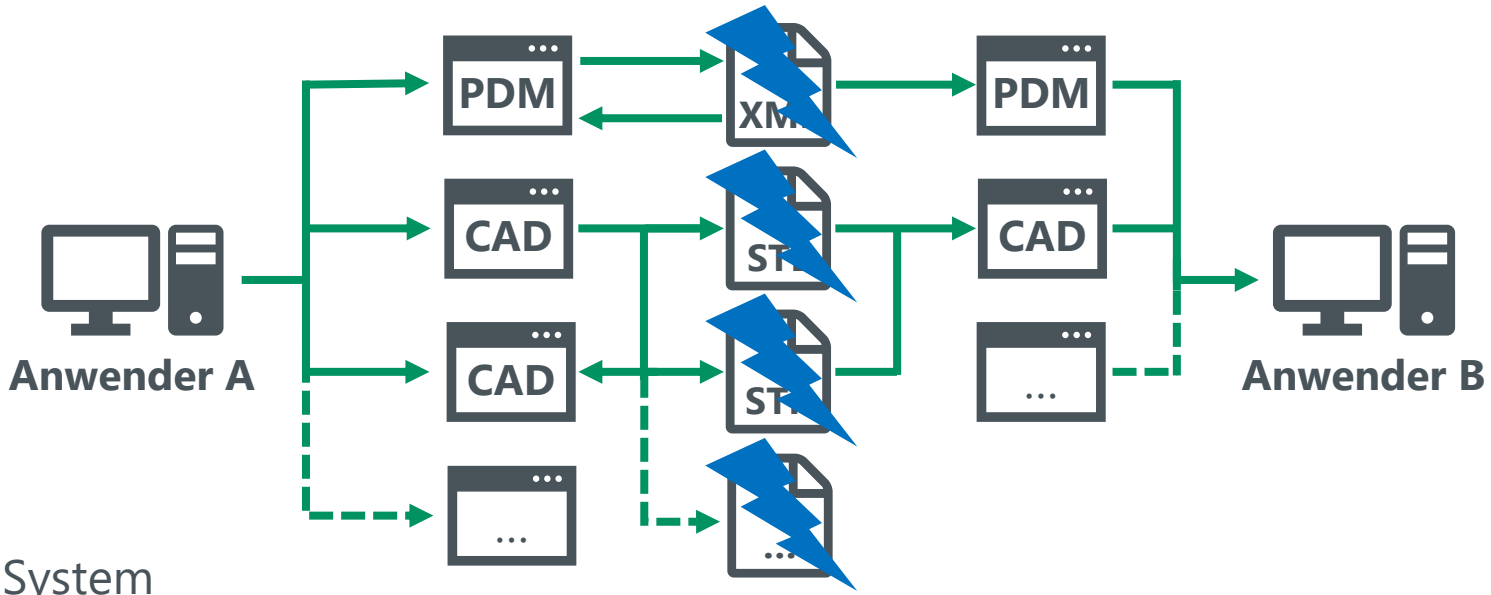
Lehrstuhl für
Konstruktionslehre und CAD
Prof. Dr.-Ing. Stephan Tremmel

Kofinanziert von der
Europäischen Union

EFRE
BAYERN 2021-2027

Motivation

Bereitstellung von
qualitativ hochwertigen Austauschdaten



Motivation

UNIVERSITÄT BAYREUTH

Wie erfolgt der Austausch von Werkzeugmodellen?

Welche Methoden eignen sich zur Datenextraktion?

Welche Besonderheiten gilt es zu beachten?

Inkompatible Werkzeug-modelle

Daten-extraktion

Daten-adaption

Kompatible Werkzeug-modelle

Wie können Daten auf Basis von regulären Ausdrücken (Regex) extrahiert werden?

Lehrstuhl für Konstruktionslehre und CAD
Prof. Dr.-Ing. Stephan Tremmel

Kofinanziert von der Europäischen Union

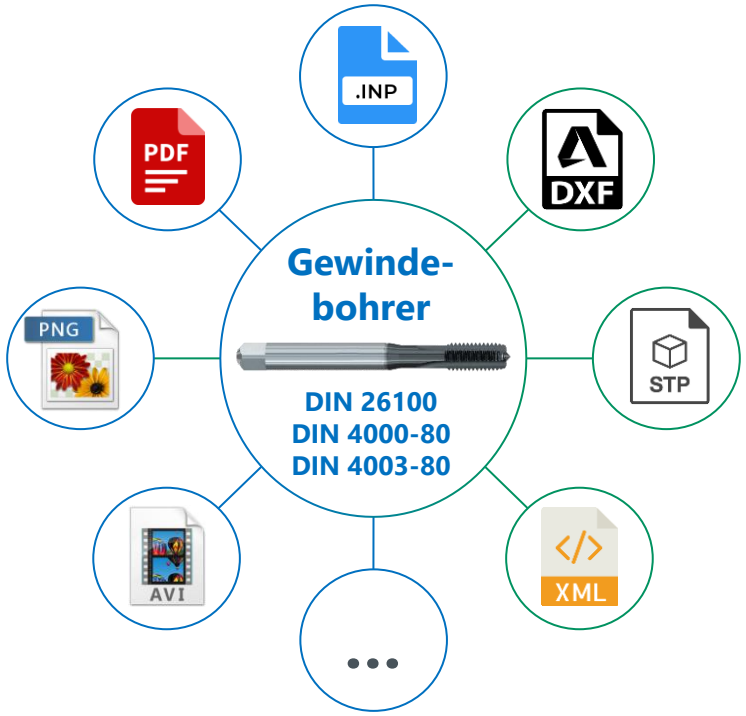
EFRE BAYERN 2021-2027

Wie erfolgt der Austausch von Werkzeugmodellen?

Aufteilung in mehrere Austauschdateien

UNIVERSITÄT BAYREUTH

Hier: Am Beispiel eines digitalen Werkzeugmodells

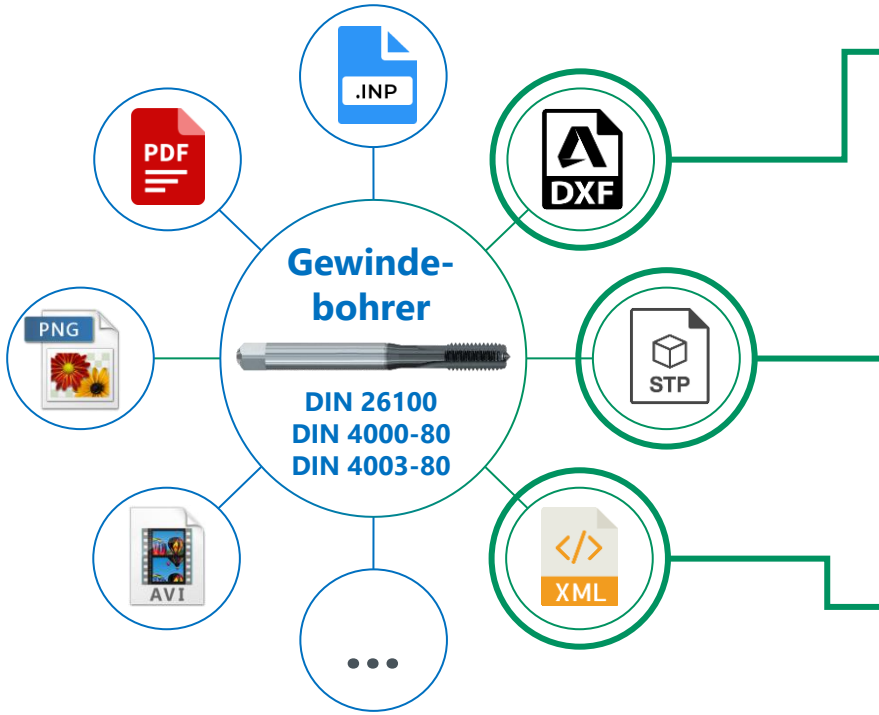


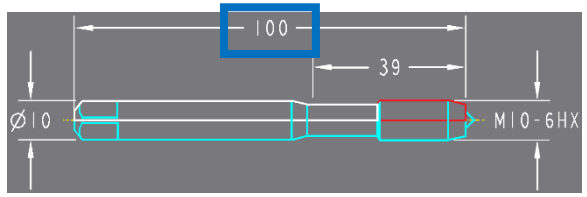
Dateiformat	Anwendungsgebiet	Gewöhnliche Kodierung
AVI	Multimedia	Binär
DXF	2D-Dokumentation, -Graphik, -Konturen	Text
PDF	2D-Dokumentation, Katalogdaten	Binär
H, I	CNC Programmierung	Text
IO, TXT, GSD	Allgemeine Beschreibungen	Binär
JPG	Skizze, Bild, Multimedia	Binär
PNG	Skizze, Foto	Binär
JT	3D-Darstellungen (detailliert und einfach)	Binär
STL; STEP	3D-Darstellungen (detailliert und einfach)	Text
NC	CNC Programmierung	Binär
P21	ISO Eigenschaften	Text
XML	ISO- und DIN-Eigenschaften, Allgemeine Beschreibungen, Katalogdaten, Dokumentation, Applikationsdaten	Text

Wie erfolgt der Austausch von Werkzeugmodellen?

Austauschdateien beinhalten eine Vielzahl an Features

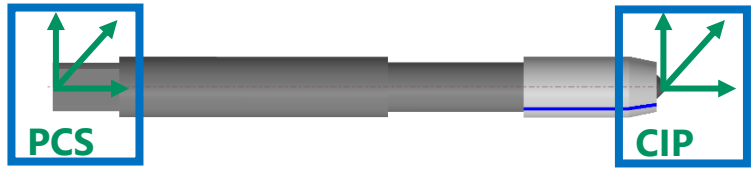
Hier: Am Beispiel eines digitalen Werkzeugmodells

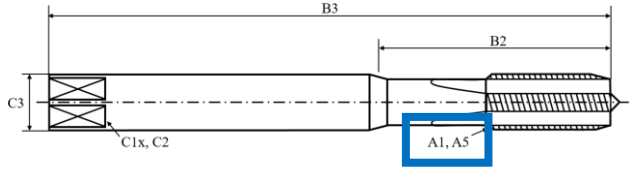




Beispiel:
Bemaßung in
einer **DXF-Datei**

Beispiel: **3D-Koordinatensystem** in einer **STEP-Datei**





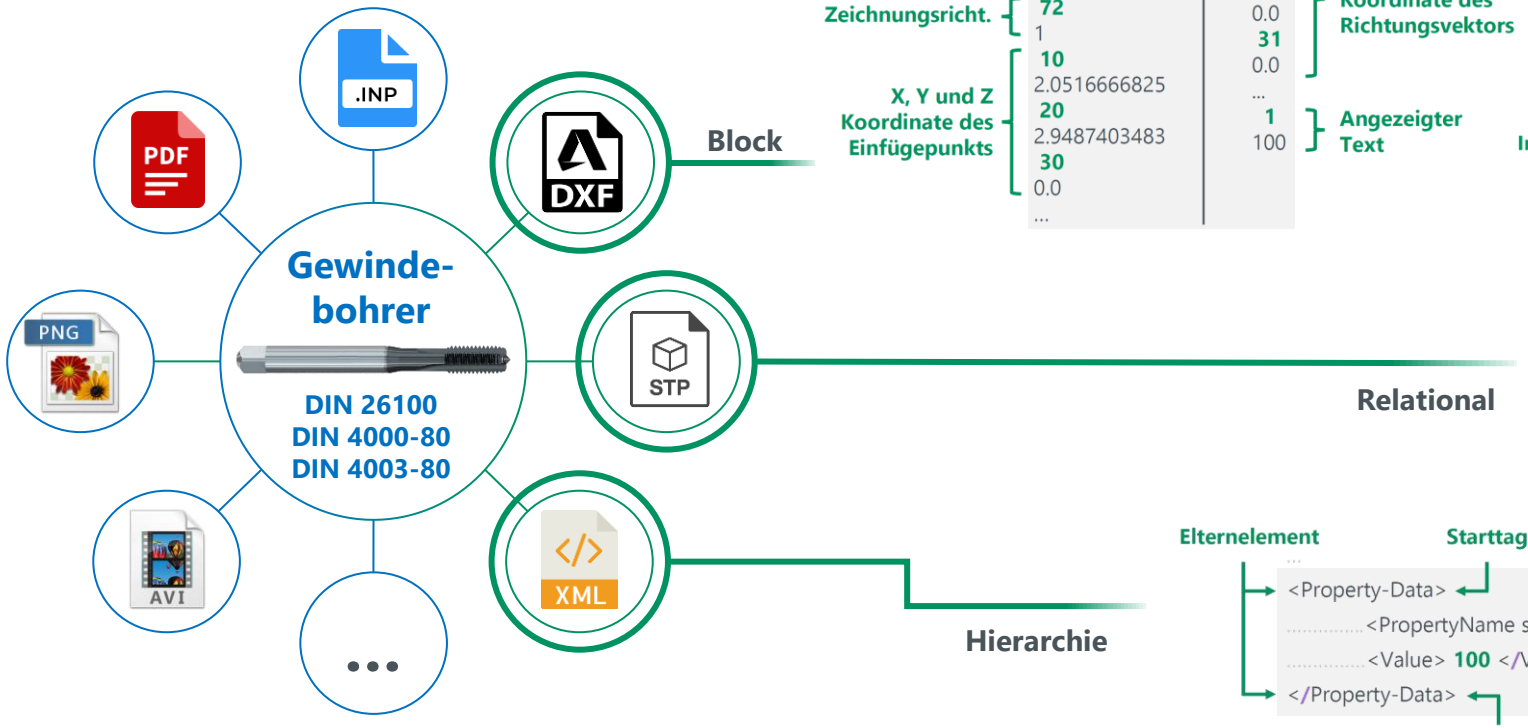
Beispiel:
Gewindetoleranzklasse
(A1) in einer **XML-Datei**

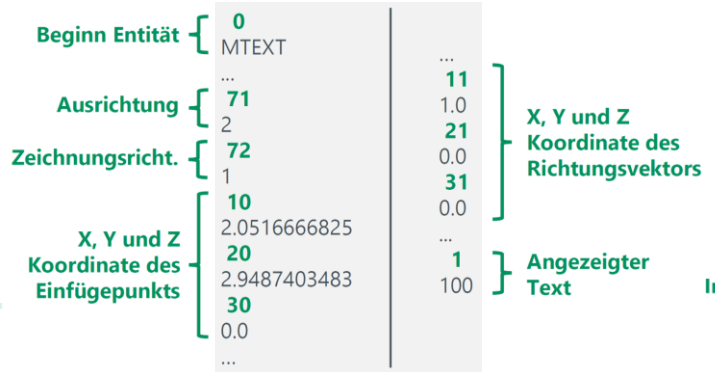
Wie erfolgt der Austausch von Werkzeugmodellen?

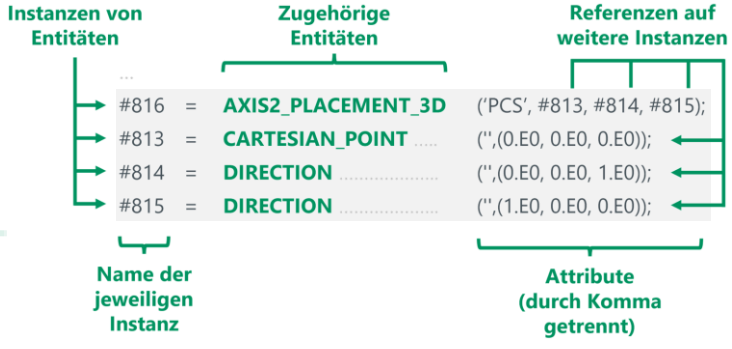
Zusammenhängende Datenstrukturen in Austauschdateien

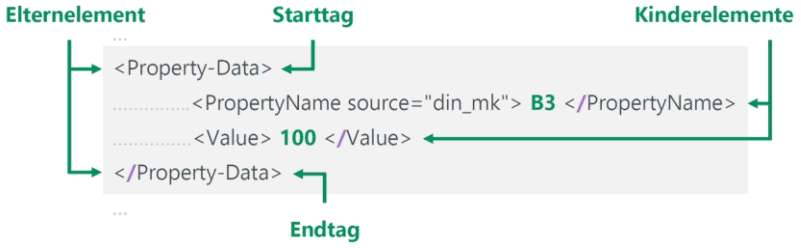
UNIVERSITÄT BAYREUTH

Hier: Am Beispiel eines digitalen Werkzeugmodells







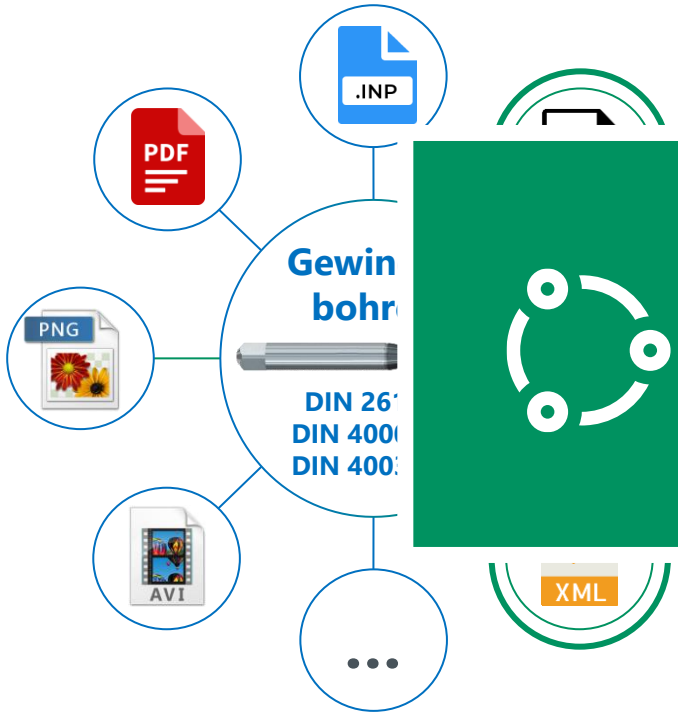


Wie erfolgt der Austausch von Werkzeugmodellen?

Zusammenhängende Datenstrukturen in Austauschdateien

UNIVERSITÄT BAYREUTH

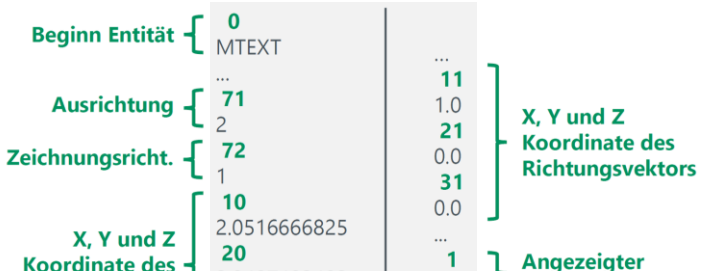
Hier: Am Beispiel eines digitalen Werkzeugmodells

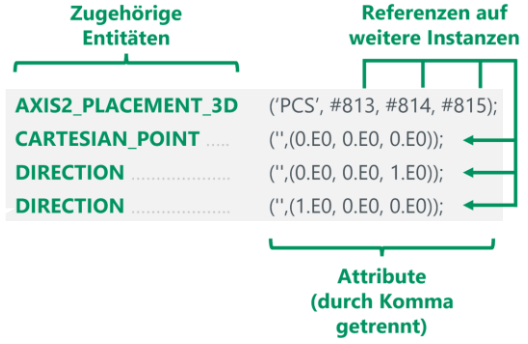


Bündelung verschiedener Daten eines Werkzeugpartialmodells

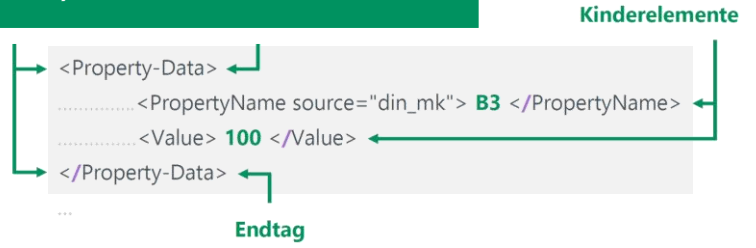
zu

logisch zusammenhängenden Einheiten mit einer höheren semantischen Bedeutung (**„Feature“**)





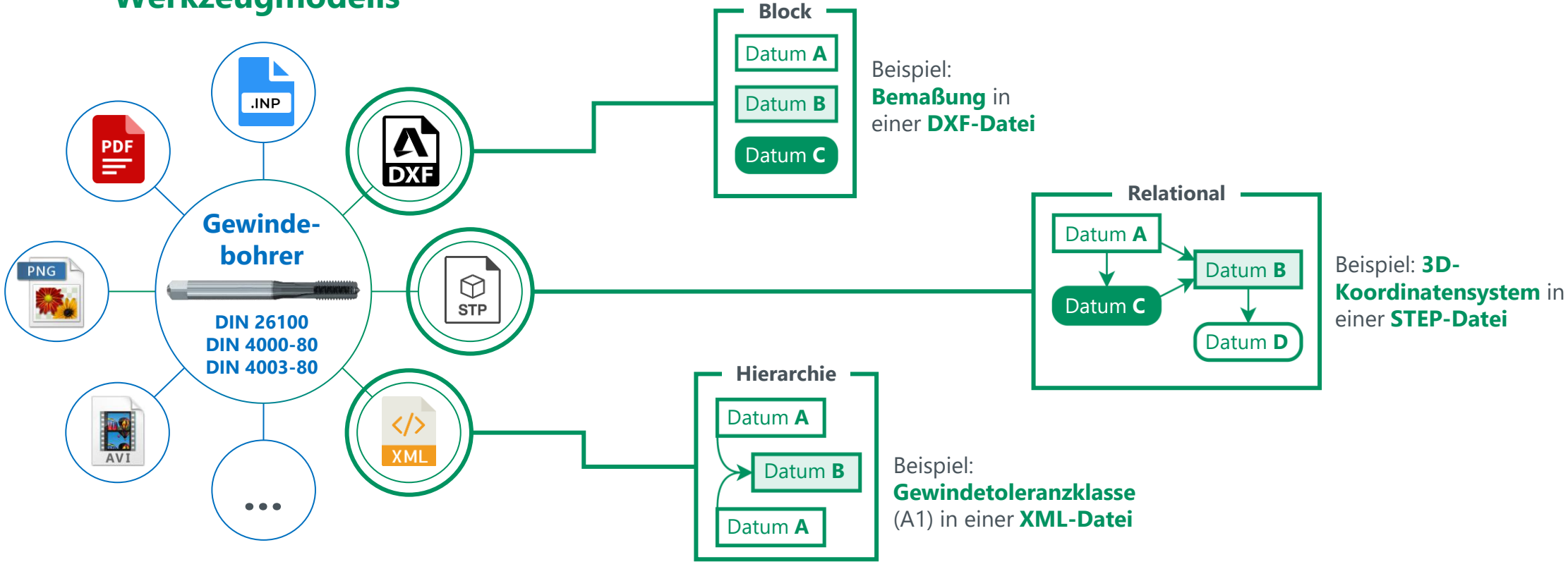
Hierarchie



Wie erfolgt der Austausch von Werkzeugmodellen?

Zusammenhängende Datenstrukturen in Austauschdateien

Hier: Am Beispiel eines digitalen Werkzeugmodells



Wie erfolgt der Austausch von Werkzeugmodellen?

Zusammenhängende Datenstrukturen in Austauschdateien

Hier: Am Beispiel eines digitalen Werkzeugmodells

Welche Methoden eignen sich zur Datenextraktion?

Akademisches Forschungsfeld der „Featureextraktion“

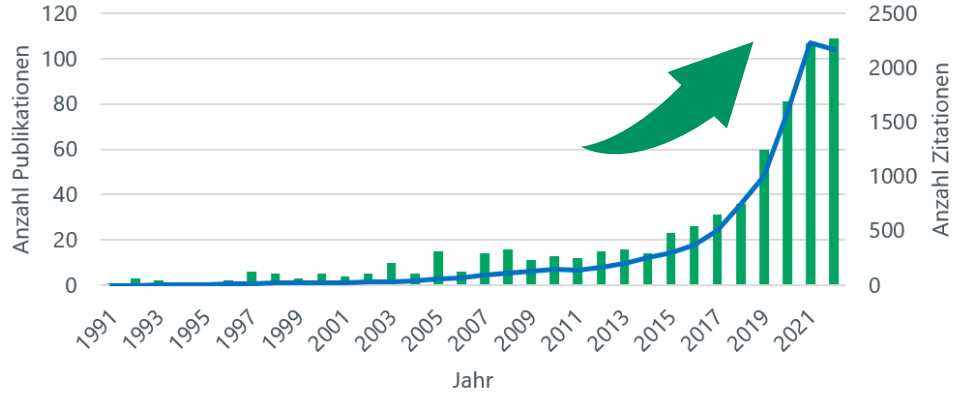
„Die Forschungsfeld der Featureextraktion umfasst den Prozess der **Identifizierung und Extraktion von relevanten Merkmalen oder Eigenschaften aus Daten**, die für eine bestimmte Aufgabe oder ein bestimmtes Modell von Bedeutung sind.“

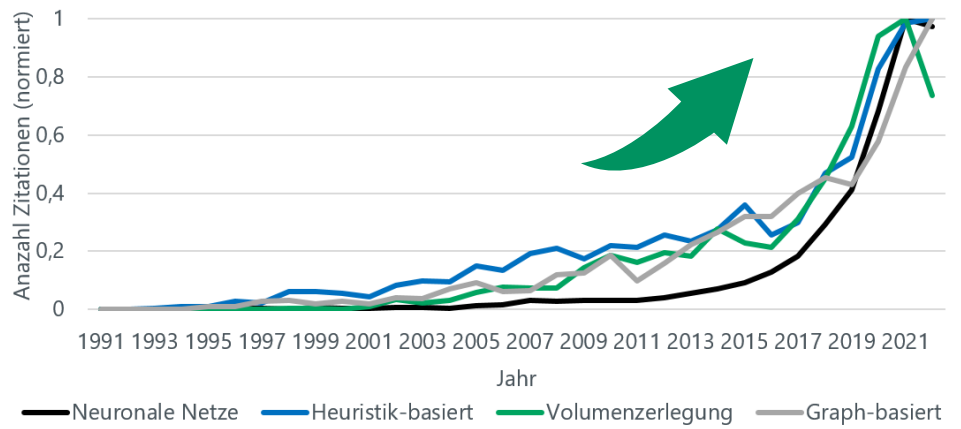
Graph-basierte Ansätze

Volumen-zerlegung

Methoden der künstlichen Intelligenz

Heuristik-basierte Ansätze





https://www.webofscience.com/wos/woscc/basic-search

Welche Methoden eignen sich zur Datenextraktion?

Warum eine Heuristik-/Hinweis-basierte Methode?

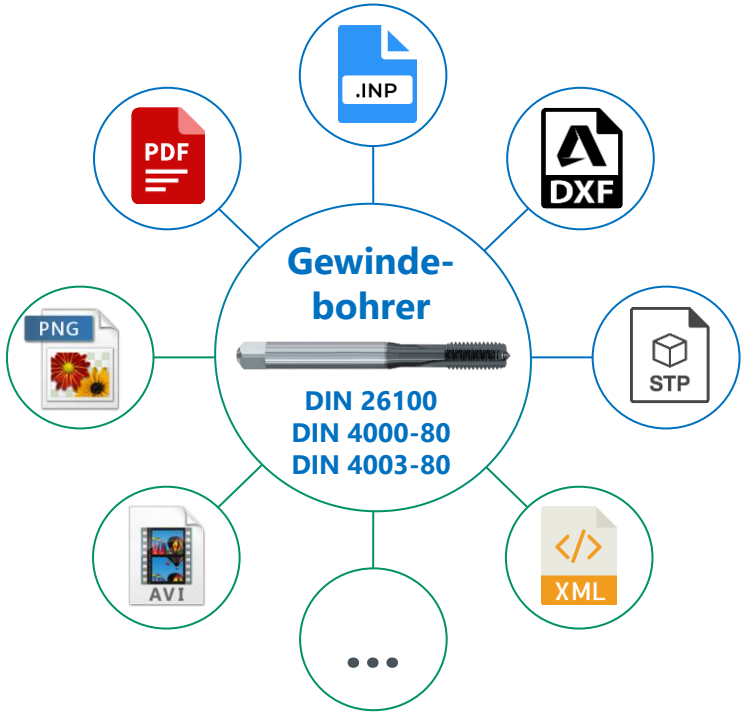
UNIVERSITÄT BAYREUTH

Vorteile Heuristik-basierter Methoden

- Gute **Nachvollziehbarkeit**
- Verhältnismäßig **unkomplizierte Anwendung**
  - ➢ Auch für **verschiedene Austauschformate**
- Identifizierung **komplexer Features** auf Basis abstrahierter Informationen (Hinweis-basiert)
- Auch **geeignet für nicht-geometrische Features**
- Möglichkeit zur Erweiterung: Hinterlegung **spezifischer Kompatibilitätsprobleme**
- **Exakte** Vorhersage
- **Metainformationen** extrahierbar (Keine Konvertierung)

Nachteile Heuristik-basierter Methoden

- **Expertenwissen** notwendig
- **Aufwendige Erstellung** für komplexe Features



Welche Methoden eignen sich zur Datenextraktion?

Warum eine Heuristik-/Hinweis-basierte Methode?

UNIVERSITÄT BAYREUTH

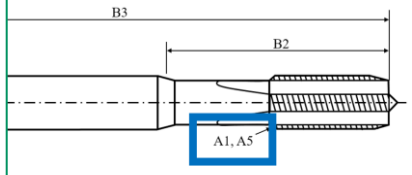
Vorteile Heuristik-basierter Methoden

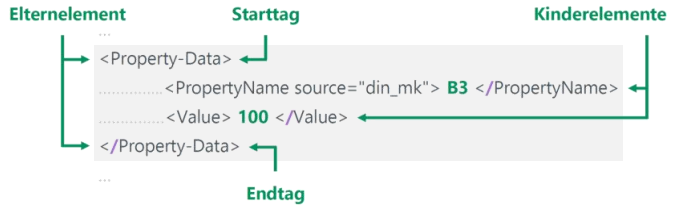
- Gute **Nachvollziehbarkeit**
- Verhältnismäßig **unkomplizierte Anwendung**
  - ➢ Auch für **verschiedene Austauschformate**
- Identifizierung **komplexer Features** auf Basis abstrahierter Informationen (Hinweis-basiert)
- Auch **geeignet für nicht-geometrische Features**
- Möglichkeit zur Erweiterung: Hinterlegung **spezifischer Kompatibilitätsprobleme**
- **Exakte** Vorhersage
- **Metainformationen** extrahierbar (Keine Konvertierung)

Nachteile Heuristik-basierter Methoden

- **Expertenwissen** notwendig
- **Aufwendige Erstellung** für komplexe Feature

? Welche Besonderheiten gilt es zu beachten?

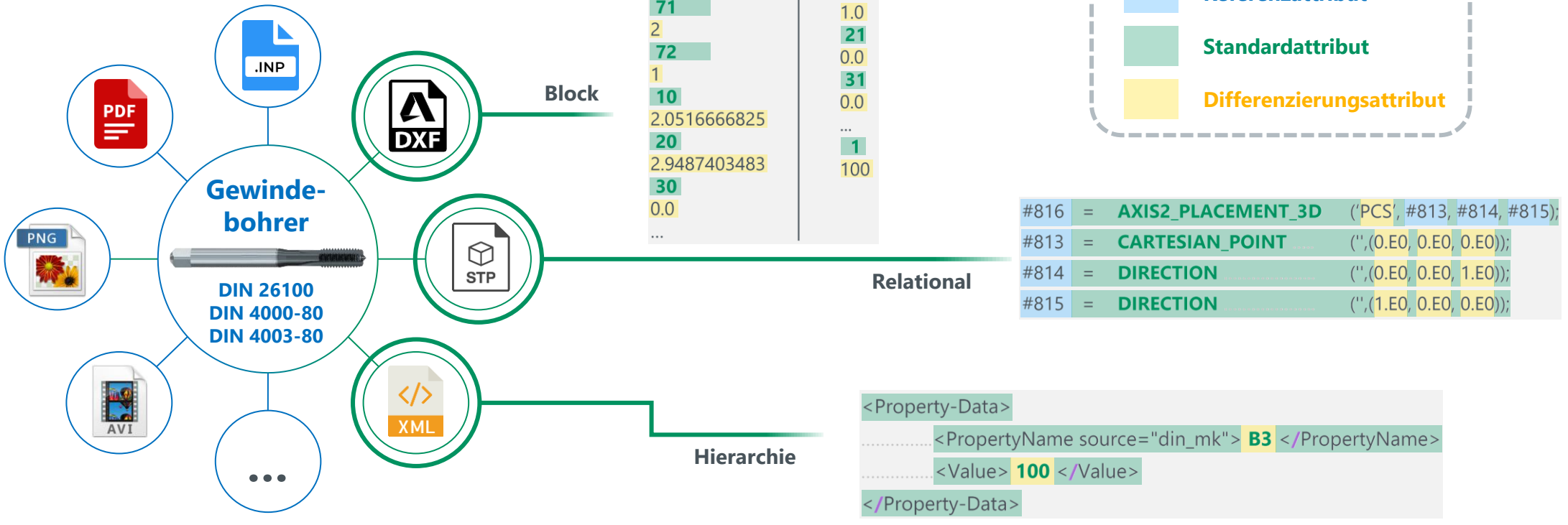




Kofinanziert von der Europäischen Union
EFRE BAYERN 2021-2027

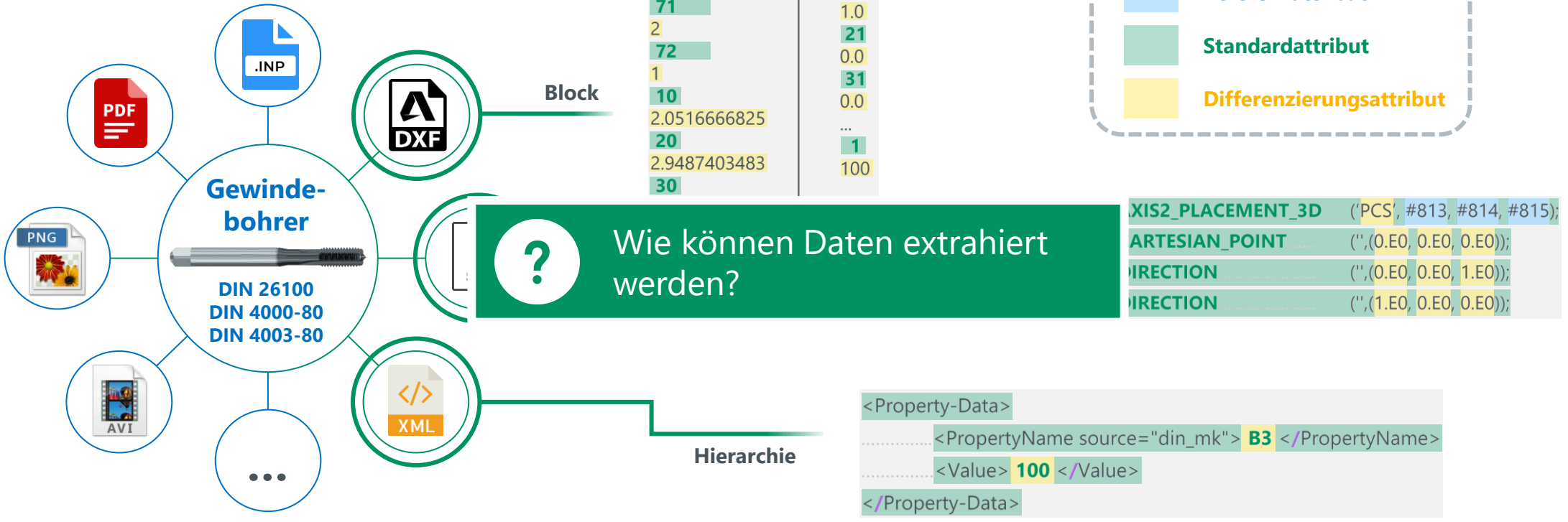
Welche Besonderheiten gilt es zu beachten?

Verschiedene Attribute



Welche Besonderheiten gilt es zu beachten?

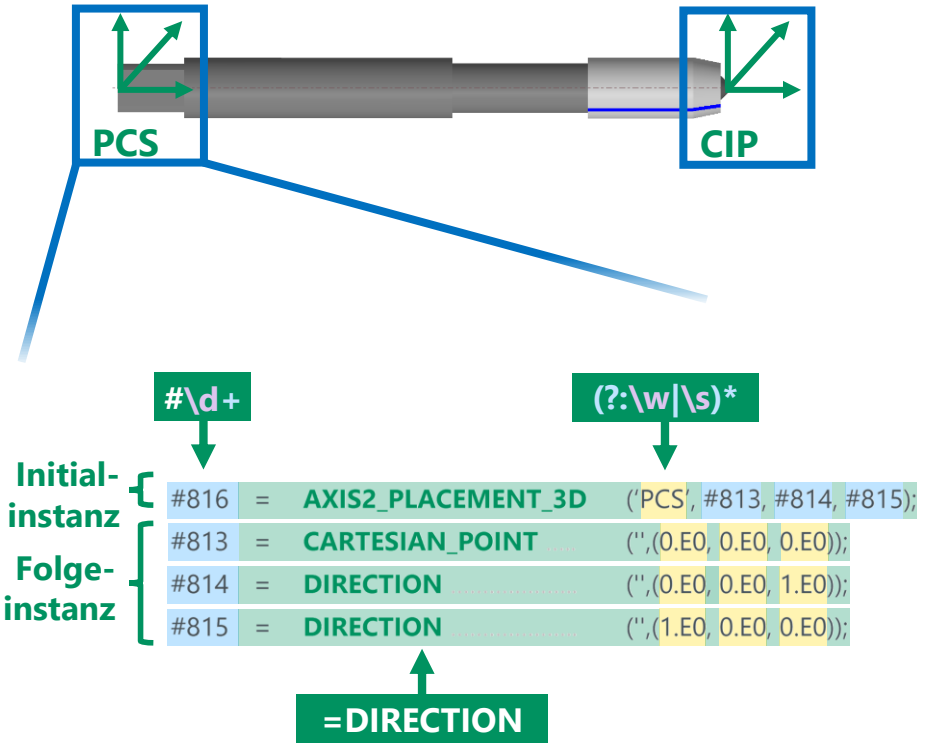
Verschiedene Attribute



Wie können Daten extrahiert werden?

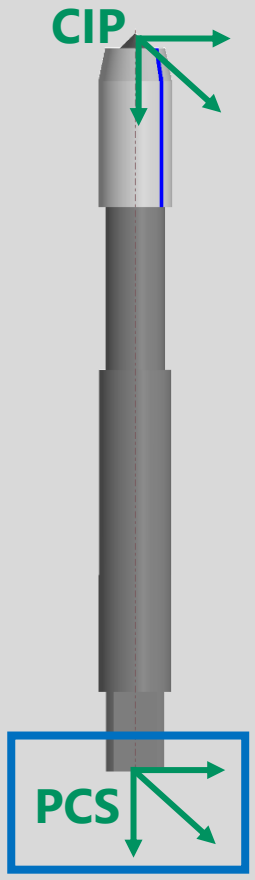
Auf Basis regulärer Ausdrücke (Regex)

- **Definition**: Suchmuster zur Textsuche und -manipulation
- **Syntax-Beispiele**:
  - `.` Jedes Zeichen
  - `*` Null oder mehr Wiederholungen
  - `|` Oder-Operator
- **Anwendungen**:
  - Textsuche
  - Datenvalidierung
  - Textersetzung
- **Unterstützte Tools**:
  - Programmiersprachen (Python, Perl, C#, etc.)
  - Texteditoren (Notepad++, Vim, etc.)



Wie können Daten extrahiert werden?

Auf Basis regulärer Ausdrücke (Regex)



WENN `#\d+=AXIS2_PLACEMENT_3D\('(?:\w|\s)+',#\d+,#\d+,#\d+\);`
matcht **DANN** Initialdaten gefunden **UND**

Matches
- `#844=AXIS2_PLACEMENT_3D('CIP',#841,#842,#843);`
- `#837=AXIS2_PLACEMENT_3D('MCS',#834,#835,#836);`
- `#816=AXIS2_PLACEMENT_3D('PCS',#813,#814,#815);`

WENN `#813=CARTESIAN_POINT\('',\(-?\d+\.(?:\d+)?[eE]-?\d+,-?\d+\.(?:\d+)?[eE]-?\d+,-?\d+\.(?:\d+)?[eE]-?\d+\)\);`
matcht **DANN** 1. Folgedatum gefunden **UND**

Match
- `#813=CARTESIAN_POINT('',(0.E0,0.E0,0.E0));`

WENN `#814=DIRECTION\('',\(-?\d+\.(?:\d+)?[eE]-?\d+,-?\d+\.(?:\d+)?[eE]-?\d+,-?\d+\.(?:\d+)?[eE]-?\d+\)\);`
matcht **DANN** 2. Folgedatum gefunden **UND**

Match
- `#814=DIRECTION('',(0.E0,0.E0,1.E0));`

WENN `#815=DIRECTION\('',\(-?\d+\.(?:\d+)?[eE]-?\d+,-?\d+\.(?:\d+)?[eE]-?\d+,-?\d+\.(?:\d+)?[eE]-?\d+\)\);`
matcht **DANN** 3. Folgedatum/ Item gefunden

Match
- `#815=DIRECTION('',(1.E0,0.E0,0.E0));`

Motivation

Zusammenfassung und Ausblick

Wie erfolgt der Austausch von Werkzeugmodellen?

Welche Methoden eignen sich zur Datenextraktion?

Welche Besonderheiten gilt es zu beachten?

Inkompatible Werkzeug-modelle

Daten-extraktion

Daten-adaption

Kompatible Werkzeug-modelle

Wie können Daten auf Basis von regulären Ausdrücken (Regex) extrahiert werden?

Lehrstuhl für Konstruktionslehre und CAD
Prof. Dr.-Ing. Stephan Tremmel

Kofinanziert von der Europäischen Union

EFRE BAYERN 2021-2027

Vielen Dank für die Aufmerksamkeit!

Quellenangabe

- Mohr, Johannes ; Kleinschrodt, Claudia: Vorgehen zur Verbesserung der Kompatiblität von Austauschdateien (GTDE Mitgliederversammlung 2022). Webkonferenz, 39.3.2022
- REINSEL, David ; RYDNING, John ; GANTZ, John: Worldwide Global DataSphere Forecast, 2021–2025: The World Keeps Creating More Data — Now, What Do We Do with It All? URL https://www.idc.com/getdoc.jsp?containerId=US46410421 – Überprüfungsdatum 2022-09-04
- Mohr, Johannes ; Kleinschrodt, Claudia ; Tremmel, Stephan ; Rieg, Frank: *Compatibility Improvement of Interrelated Items in Exchange Files—A General Method for Supporting the Data Integrity of Digital Twins*. In: *Applied Sciences* 12 (2022), Nr. 16, S. 8099
- DIN Deutsches Institut für Normung: *DIN 4000-80 : Sachmerkmal-Listen - Teil 80: Gewindefurcher und Schneideisen*. Berlin : Beuth, 2019
- DIN Deutsches Institut für Normung: *DIN 4003-80 : Konzept für den Aufbau von 3D-Modellen auf Grundlage von Merkmalen nach DIN 4000 - Teil 80: Gewindebohrer, Gewindefurcher und Schneideisen*. Berlin : Beuth, 2019
- DIN Deutsches Institut für Normung: *DIN 26100 : Container-Datei - Zusammenfassung verschiedener Produktdateien für den Datenaustausch*. Berlin : Beuth, 2021
- Mohr, Johannes ; Tremmel, Stephan ; Rieg, Frank: Methoden der Featureextraktion und deren Potentiale zur Kompatibilitätsverbesserung von Austauschdateien (VDMA Technologieforum 2022). Stuttgart, 15.9.2022
- Highly Efficient and Scalable Technique for Matching Regex Patterns. In: Association for Computing Machinery (Hrsg.): Proceedings of the 2018 2nd High Performance Computing and Cluster Technologies Conference. New York, NY, USA : ACM, 2018, S. 69–78
- Nagy, Zsolt: Regex Quick Syntax Reference : Understanding and Using. Berkeley, CA : Apress, 2018
- Kleinschrodt, Claudia ; Mohr, Johannes ; Zimmermann, Markus ; Rieg, Frank: Konzeptionelles Design zur softwaregestützten Analyse und Modifikation von Produktdaten, Bd. 16. In: Brökel, Klaus; Nagarajah, Arun; Rieg, Frank; Scharr, Gerhard; Stelzer, Ralph (Hrsg.): Digitalisierung und Produktentwicklung. Bayreuth : Universität Bayreuth, 2018, S. 156–167
- Corves, Burkhard (Hrsg.); Gericke, Kilian (Hrsg.); Grote, Karl-Heinrich (Hrsg.); Lohrengel, Armin (Hrsg.); Müller, Norbert (Hrsg.); Nagarajah, Arun (Hrsg.); Rieg, Frank (Hrsg.); Scharr, Gerhard (Hrsg.); Stelzer, Ralph (Hrsg.): 17. Gemeinsames Kolloquium Konstruktionstechnik: Agile Entwicklung physischer Produkte, 2019