

XML in KNIME

Basics

What is XML?

- eXtensible Markup Language using “tags”
- Human and machine readable text file
- Data exchange standard between machines in the web
- Software independent
- Data with rules & structure
 - Descriptive markup vs. procedural and presentational
- HTML is used to display data
 - predefined tags
- XML is used to store data
 - no predefined tags

How does XML look like?

<Prolog: Declaration>

<?xml version = "1.0" encoding = "UTF-8"?>

<RootElement>

<Qualitaetsbericht>

 <Element1>

 <Datensatz>

 <Elem1.1>data</Elem1.1>

 <Datum>2014-07-05</Datum>

 <Elem1.2>data</Elem1.2>

 <Uhrzeit>00:27:24</Uhrzeit>

 <Elem1.3>data</Elem1.3>

 <Sprache>deutsch</Sprache>

 </Element1>

 </Datensatz>

</RootElement>

</Qualitaetsbericht>

XML Lingo

XML declaration

Root element starting tag

Parent element starting tag

Child element starting /ending tag

Sibling element starting /ending tag

/Parent element ending tag

/Root element ending tag

XPath - the XML Path Language

- XPath is a query language for XML files
- Search for ELEMENTS
- Retrieve DATA or child elements
- Search expression is called **Location Path**

Location Path for <Sprache>

/Qualitaetsbericht/Datensatz/Sprache

Query Result: **“deutsch”**

```
<?xml version = "1.0" encoding = "UTF-8"?>
```

```
<Qualitaetsbericht>
```

```
<Datensatz>
```

```
<Datum>2014-07-05</Datum>
```

```
<Uhrzeit>00:27:24</Uhrzeit>
```

```
<Sprache>deutsch</Sprache>
```

```
</Datensatz>
```

```
</Qualitaetsbericht>
```

Using XPath to Create Data Tables

- Sibling element tags define column headers
- ONE parent element is ONE row

//Hauptdiagnosen/Hauptdiagnose/ICD_10

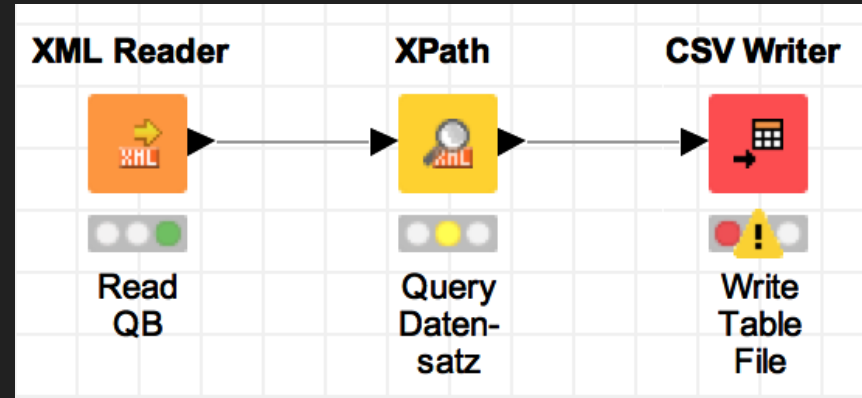
//Hauptdiagnosen/Hauptdiagnose/Fallzahl

ICD-10	Fallzahl
I48.1	444
I50.1	332
R55	295
I20.0	265

```
...
<Hauptdiagnosen>
  <Hauptdiagnose>
    <ICD_10>I48.1</ICD_10>
    <Fallzahl>444</Fallzahl>
  </Hauptdiagnose>
  <Hauptdiagnose>
    <ICD_10>I50.1</ICD_10>
    <Fallzahl>332</Fallzahl>
  </Hauptdiagnose>
  <Hauptdiagnose>
    <ICD_10>R55</ICD_10>
    <Fallzahl>295</Fallzahl>
  </Hauptdiagnose>
  <Hauptdiagnose>
    <ICD_10>I20.0</ICD_10>
    <Fallzahl>265</Fallzahl>
  </Hauptdiagnose>
</Hauptdiagnosen>
...
```

What is KNIME?

- Open Source platform for
 - data analytics
 - predictive analytics
 - modelling
- KNIME has a graphical interface instead of a scripting language like R
- You use KNIME by creating a workflow consisting of nodes
- Drag&Drop nodes from node repository onto canvas
- Nodes are connected and configured
- Run the workflow (step by step)
- Traffic light indicator
 - Red = cannot run
 - Yellow = ready to run
 - Green = successful run



Qualitätsbericht Structure

XPath to ICD10 codes:

Qualitätsbericht /
Organisationseinheiten_Fachabteilungen /
Organisationseinheiten_Fachabteilung /
Hauptdiagnosen

...need to extract IK and Gliederungsnummer to
match data!

How to combine data from different XML hierarchy levels?

```
<?xml version="1.0" encoding="UTF-8"?>
<Qualitaetsbericht>
  <Datensatz> ☐ ☐ </Datensatz>
  <Producer>promato QB</Producer>
  <Einleitung> ☐ ☐ </Einleitung>
  <Krankenhaus>
    <Kontaktdaten>
      <Name>DRK-Krankenhaus Mölln-Ratzeburg gGmbH</Name>
      <IK>260100454</IK>
      <Standortnummer>00</Standortnummer>
      <Hausanschrift> ☐ ☐ </Hausanschrift>
      <Postanschrift> ☐ ☐ </Postanschrift>
      <Telefon> ☐ ☐ </Telefon>
      <Fax> ☐ ☐ </Fax>
      <Email>info@drk-krankenhaus.de</Email>
      <URL>http://www.drk-krankenhaus.de</URL>
      <Aerztliche_Leitung> ☐ ☐ </Aerztliche_Leitung>
      <Pflegedienstleitung> ☐ ☐ </Pflegedienstleitung>
      <Verwaltungsleitung> ☐ ☐ </Verwaltungsleitung>
    </Kontaktdaten>
  </Krankenhaus>
  <Einziger_Standort/>
  <Krankenhaustraeger> ☐ ☐ </Krankenhaustraeger>
  <Akademisches_Lehrkrankenhaus> ☐ ☐ </Akademisches_Lehrkrankenhaus>
  <Psychiatrisches_Krankenhaus/>
  <Anzahl_Betten>184</Anzahl_Betten>
  <Fallzahlen> ☐ ☐ </Fallzahlen>
  <Medizinisch_PflegerischeLeistungsangebote> ☐ ☐ </Medizinisch_PflegerischeLeistungsangebote>
  <Nicht_MedizinischeLeistungsangebote> ☐ ☐ </Nicht_MedizinischeLeistungsangebote>
  <Barrierefreiheit> ☐ ☐ </Barrierefreiheit>
  <Akademische_Lehre> ☐ ☐ </Akademische_Lehre>
  <Ausbildung_andere_Heilberufe> ☐ ☐ </Ausbildung_andere_Heilberufe>
  <Personal_des_Krankenhauses> ☐ ☐ </Personal_des_Krankenhauses>
  <Verantwortliche_Qualitaetsmanagement> ☐ ☐ </Verantwortliche_Qualitaetsmanagement>
  <Apparative_Ausstattung> ☐ ☐ </Apparative_Ausstattung>
  <Beschwerdemanagement> ☐ ☐ </Beschwerdemanagement>
  <Organisationseinheiten_Fachabteilungen>
    <Organisationseinheit_Fachabteilung> ☐ ☐ </Organisationseinheit_Fachabteilung>
    <Organisationseinheit_Fachabteilung>
      <Gliederungsnummer>2</Gliederungsnummer>
      <Name>Allgemeine Chirurgie</Name>
      <Chefarzt>Dr. med. MaHM Andreas Schmid</Chefarzt>
      <Chefarzt>PD Dr. Peter Benecke</Chefarzt>
      <Kontaktdaten> ☐ ☐ </Kontaktdaten>
      <Zielvereinbarungen> ☐ ☐ </Zielvereinbarungen>
      <Ambulante_D_Arzt_Zulassung/>
      <Stationaere_BG_Zulassung/>
      <Fachabteilungsschlüssel>
        <FA_Schlüssel>1500</FA_Schlüssel>
      </Fachabteilungsschlüssel>
      <MedizinischeLeistungsangebote> ☐ ☐ </MedizinischeLeistungsangebote>
      <Barrierefreiheit> ☐ ☐ </Barrierefreiheit>
      <Fallzahlen_OE> ☐ ☐ </Fallzahlen_OE>
      <Hauptdiagnosen> ☐ ☐ </Hauptdiagnosen>
      <Prozeduren> ☐ ☐ </Prozeduren>
      <Ambulante_Behandlungsmoeglichkeiten> ☐ ☐ </Ambulante_Behandlungsmoeglichkeiten>
      <Ambulante_Operationen> ☐ ☐ </Ambulante_Operationen>
      <Personelle_Ausstattung> ☐ ☐ </Personelle_Ausstattung>
    </Organisationseinheit_Fachabteilung>
    <Organisationseinheit_Fachabteilung> ☐ ☐ </Organisationseinheit_Fachabteilung>
    <Organisationseinheit_Fachabteilung> ☐ ☐ </Organisationseinheit_Fachabteilung>
    <Organisationseinheit_Fachabteilung> ☐ ☐ </Organisationseinheit_Fachabteilung>
  </Organisationseinheiten_Fachabteilungen>
  <Qualitaetssicherung> ☐ ☐ </Qualitaetssicherung>
</Qualitaetsbericht>
```

How should the data table look like?

- Redundant categorical info (e.g. for filtering)
- Attribute Value pairs

In KNIME this table is developed step by step out of the XML and several XPath queries.

IK	Gliederungsnummer	ICD_10	Fallzahl
260100454	2	M84.4	
260100454	2	M84.5	8
260100454	2	M84.6	4
260100454	2	M84.7	6
260100454	2	M84.8	
260100454	2	M84.9	7
260100454	2	M84.10	5
260100454	2	M84.11	10
260100454	2	M84.12	4
260100454	2	M84.13	7

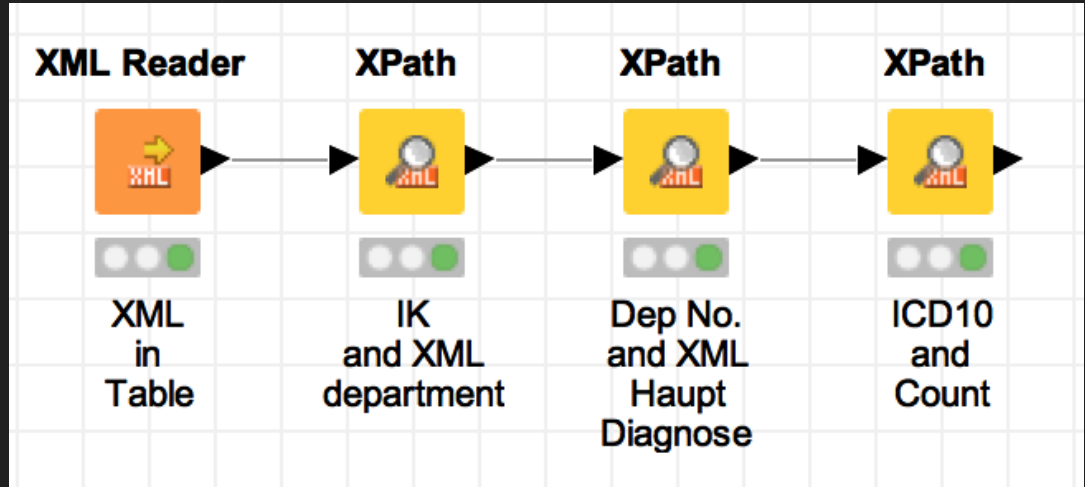
KNIME XPath sequence

XML Reader: loads the entire file

XPath 1: extracts IK number and department section as multiple XMLs

XPath 2: extracts department number and extracts diagnosis section as multiple XML

XPath 3: extracts ICD10 codes and the corresponding count (Fallzahl)



Zooming into the XML structure

XPath zoom into XML

1. IK number as string

- a. XML of all departments

2. Department number

- a. XML of all diagnoses

3. All diagnoses as

- a. ICD10
- b. Count

1.

2.

3.

```
<?xml version="1.0" encoding="UTF-8"?>
<Qualitaetsbericht>
  <Datensatz> ☐ ☐ /Datensatz>
  <Producer>promato QB</Producer>
  <Einleitung> ☐ ☐ /Einleitung>
  <Krankenhaus>
    <Kontaktdaten>
      <Name>DRK-Krankenhaus Mölln-Ratzeburg gGmbH</Name>
      <IK>260100454</IK>
      <Standortnummer>00</Standortnummer>
      <Hausanschrift> ☐ ☐ /Hausanschrift>
      <Postanschrift> ☐ ☐ /Postanschrift>
      <Telefon> ☐ ☐ /Telefon>
      <Fax> ☐ ☐ /Fax>
      <Email>info@drk-krankenhaus.de</Email>
      <URL>http://www.drk-krankenhaus.de</URL>
      <Aerztliche_Leitung> ☐ ☐ /Aerztliche_Leitung>
      <Pflegedienstleitung> ☐ ☐ /Pflegedienstleitung>
      <Verwaltungsleitung> ☐ ☐ /Verwaltungsleitung>
    </Kontaktdaten>
  </Krankenhaus>
  <Einziger_Standort/>
  <Krankenhaustraeger> ☐ ☐ /Krankenhaustraeger>
  <Akademisches_Lehrkrankenhaus> ☐ ☐ /Akademisches_Lehrkrankenhaus>
  <Psychiatrisches_Krankenhaus/>
  <Anzahl_Betten>184</Anzahl_Betten>
  <Fallzahlen> ☐ ☐ /Fallzahlen>
  <Medizinisch_PflegerischeLeistungsangebote> ☐ ☐ /Medizinisch_PflegerischeLeistungsangebote>
  <Nicht_MedizinischeLeistungsangebote> ☐ ☐ /Nicht_MedizinischeLeistungsangebote>
  <Barrierefreiheit> ☐ ☐ /Barrierefreiheit>
  <Akademische_Lehre> ☐ ☐ /Akademische_Lehre>
  <Ausbildung_andere_Heilberufe> ☐ ☐ /Ausbildung_andere_Heilberufe>
  <Personal_des_Krankenhauses> ☐ ☐ /Personal_des_Krankenhauses>
  <Verantwortliche_Qualitaetsmanagement> ☐ ☐ /Verantwortliche_Qualitaetsmanagement>
  <Apparative_Ausstattung> ☐ ☐ /Apparative_Ausstattung>
  <Beschwerdemanagement> ☐ ☐ /Beschwerdemanagement>
  <Organisationseinheiten_Fachabteilungen>
    <Organisationseinheit_Fachabteilung> ☐ ☐ /Organisationseinheit_Fachabteilung>
    <Organisationseinheit_Fachabteilung>
      <Gliederungsnummer>2</Gliederungsnummer>
      <Name>Allgemeine Chirurgie</Name>
      <Chefarzt>Dr. med. MaHM Andreas Schmid</Chefarzt>
      <Chefarzt>PD Dr. Peter Benecke</Chefarzt>
      <Kontaktdaten> ☐ ☐ /Kontaktdaten>
      <Zielvereinbarungen> ☐ ☐ /Zielvereinbarungen>
      <Ambulante_D_Arzt_Zulassung/>
      <Stationaere_BG_Zulassung/>
      <Fachabteilungsschlüssel>
        <FA_Schlüssel>1500</FA_Schlüssel>
      </Fachabteilungsschlüssel>
      <MedizinischeLeistungsangebote> ☐ ☐ /MedizinischeLeistungsangebote>
      <Barrierefreiheit> ☐ ☐ /Barrierefreiheit>
      <Fallzahlen_OE> ☐ ☐ /Fallzahlen_OE>
      <Hauptdiagnosen> ☐ ☐ /Hauptdiagnosen>
      <Prozeduren> ☐ ☐ /Prozeduren>
      <Ambulante_Behandlungsmoeglichkeiten> ☐ ☐ /Ambulante_Behandlungsmoeglichkeiten>
      <Ambulante_Operationen> ☐ ☐ /Ambulante_Operationen>
      <Personelle_Ausstattung> ☐ ☐ /Personelle_Ausstattung>
    </Organisationseinheit_Fachabteilung>
    <Organisationseinheit_Fachabteilung> ☐ ☐ /Organisationseinheit_Fachabteilung>
    <Organisationseinheit_Fachabteilung> ☐ ☐ /Organisationseinheit_Fachabteilung>
    <Organisationseinheit_Fachabteilung> ☐ ☐ /Organisationseinheit_Fachabteilung>
    <Organisationseinheit_Fachabteilung> ☐ ☐ /Organisationseinheit_Fachabteilung>
  </Organisationseinheiten_Fachabteilungen>
  <Qualitaetssicherung> ☐ ☐ /Qualitaetssicherung>
</Qualitaetsbericht>
```