

Integrierte lexikographische Dienste zur Unterstützung der digitalen Geisteswissenschaften

**Karlheinz Mörth
Hannes Pirker**

***Austrian Center for Digital Humanities*
Österreichischen Akademie der Wissenschaften
(ACDH-ÖAW)**

Baffle

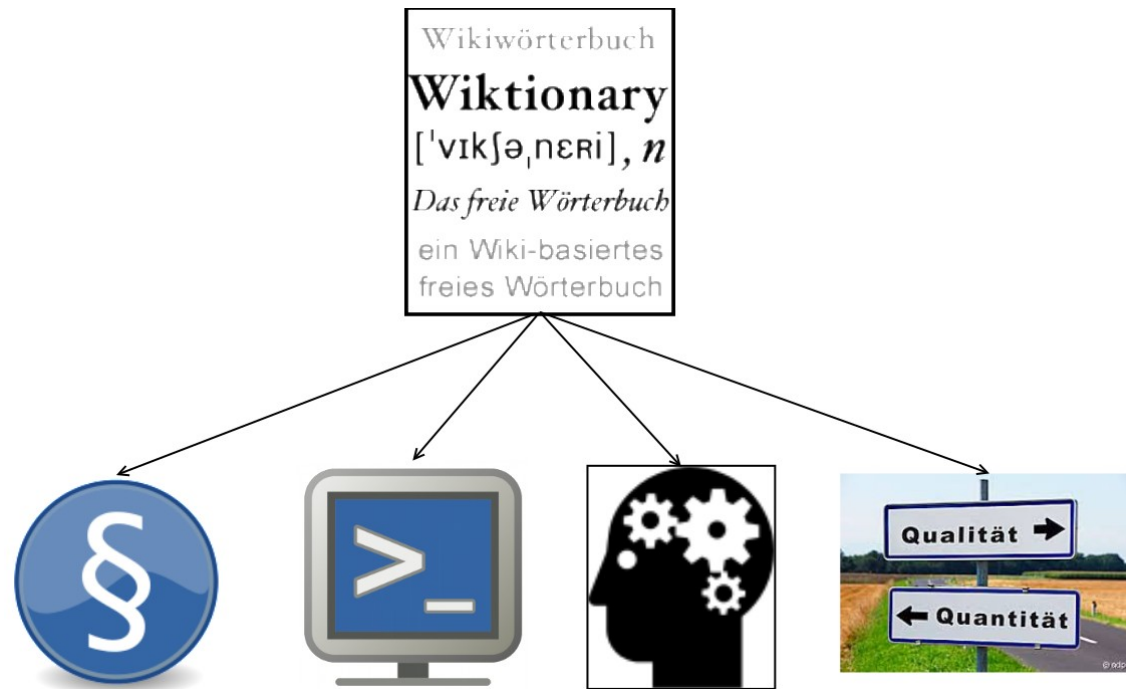
basic Austrian fullform lexicon

- *basic* im Sinne von „nutzbar als Basis“
 - freie Verfügbarkeit
 - verarbeitungsgerechtes Format
 - validierte Inhalte
- *basic* im Sinne von „beschränkter Inhalt“
 - morphologische Information
 - Frequenzdaten
- Methodik
 - korpusgestützte semi-automatische Lexikonerweiterung
 - Ausgangspunkt: Wiktionary

Wiktionary

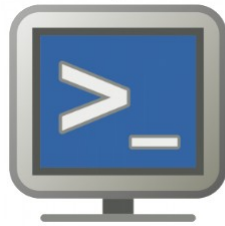
- Freies Wörterbuch
- ca. 75 000 Einträge (Deutsch)
- Autoren: „lexikalische Enthusiasten“
- mehrsprachig (Englisch, Französisch, Polnisch, Vietnamesisch)

Wiktionary



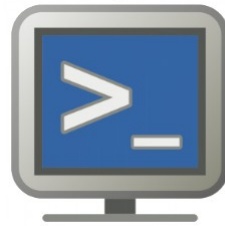
Wiktionary

Wikiwörterbuch
Wiktionary
[ˈvɪkʃəˌnɛʀi], *n*
Das freie Wörterbuch
ein Wiki-basiertes
freies Wörterbuch

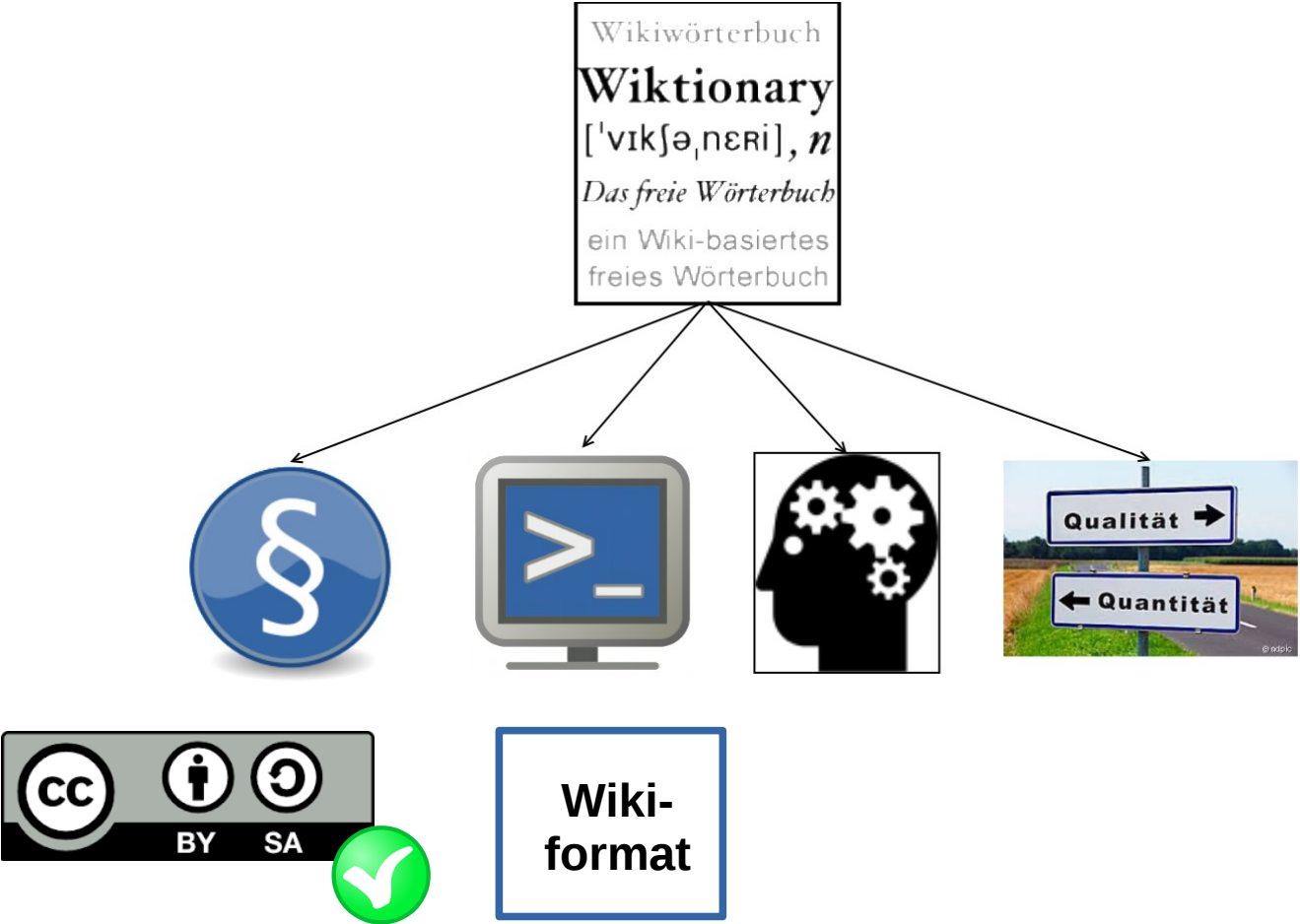


Wiktionary

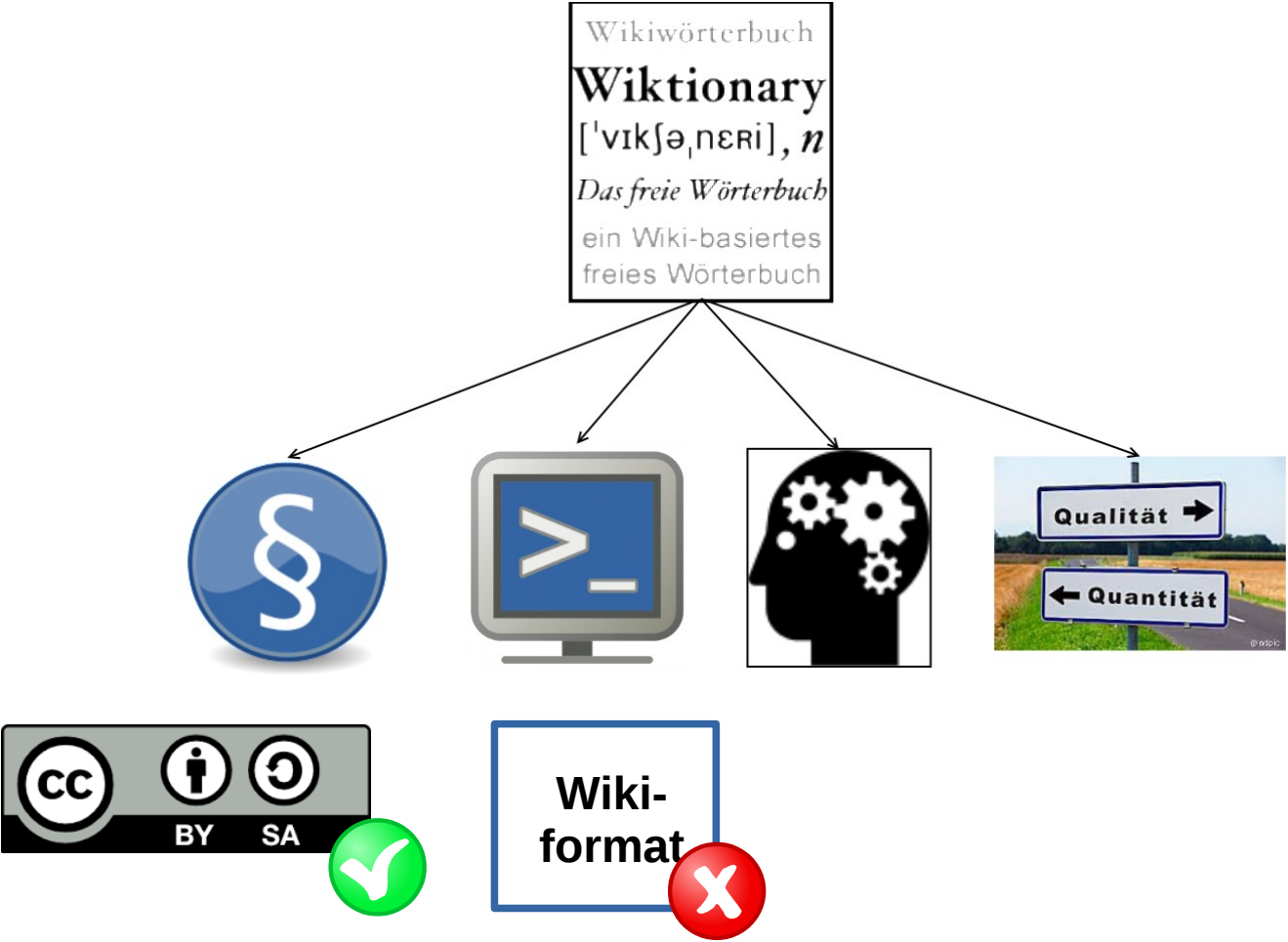
Wikiwörterbuch
Wiktionary
[ˈvɪkʃəˌnɛʀi], *n*
Das freie Wörterbuch
ein Wiki-basiertes
freies Wörterbuch



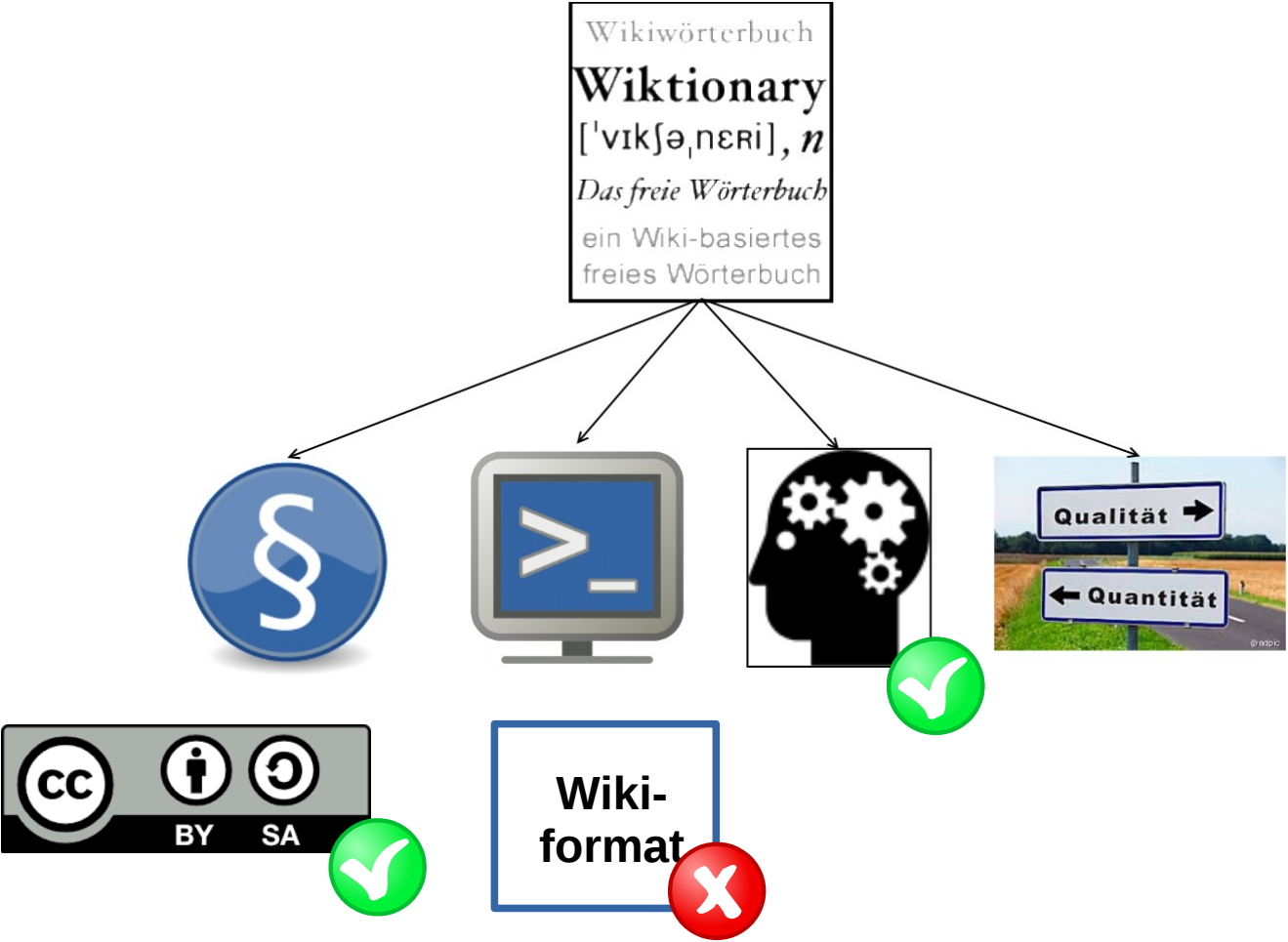
Wiktionary



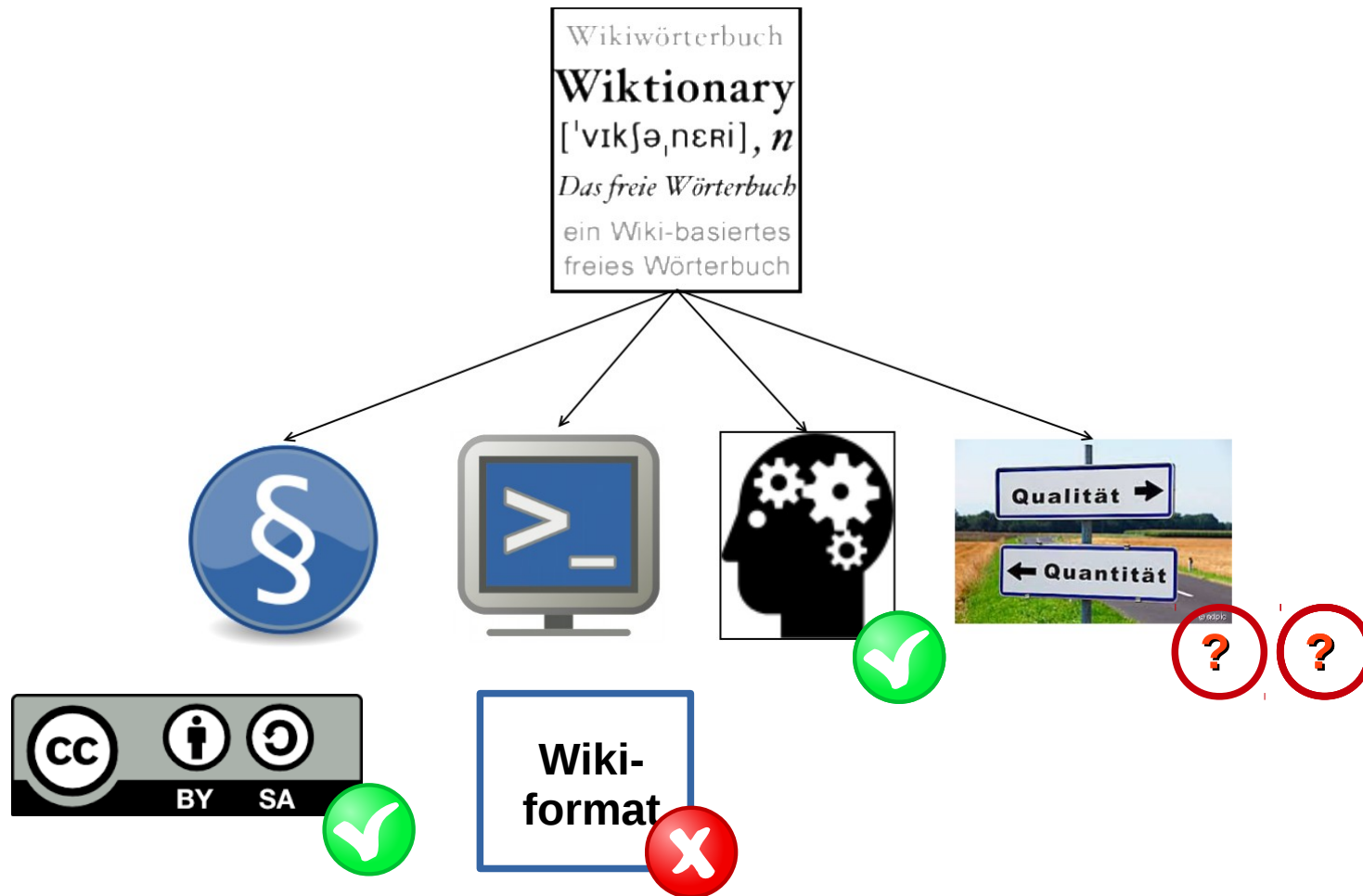
Wiktionary



Wiktionary



Wiktionary



Wiktionary: Beispielertrag

Wörterbuch (Deutsch) [\[Bearbeiten\]](#)

Substantiv, n [\[Bearbeiten\]](#)

Worttrennung:

Wör-ter-buch, Plural: Wör-ter-bü-cher

Aussprache:

IPA: [ˈvœɐ̯tɐ, buːχ], Plural: [ˈvœɐ̯tɐ, byːçɐ]

Hörbeispiele:  Wörterbuch

(österreichisch) [\(Info\)](#), Plural:  Wörterbücher

(österreichisch) [\(Info\)](#)

Bedeutungen:

[1] Nachschlagewerk für die Schreibweise, Bedeutung, Grammatik, Geschichte und/oder Übersetzung von Wörtern

Abkürzungen:

[1] **Wb.**, **Wtb.**

Herkunft:

Verdeutschung von *dictionarium* bzw. *lexicon* im 17. Jahrhundert durch [Comenius](#)^[1], ursprünglich holländisch: *wordboek*.

Synonyme:

[1] *manchmal auch:* [Diktionär](#), [Enzyklopädie](#), [Lexikon](#), [Thesaurus](#)

	Singular	Plural
Nominativ	das Wörterbuch	die Wörterbücher
Genitiv	des Wörterbuchs des Wörterbuches	der Wörterbücher
Dativ	dem Wörterbuch dem Wörterbuche	den Wörterbüchern
Akkusativ	das Wörterbuch	die Wörterbücher



[1] Lateinisches Wörterbuch 

Wiktionary: Beispieleintrag

Wörterbuch (Deutsch) [\[Bearbeiten\]](#)

Substantiv, n [\[Bearbeiten\]](#)

Worttrennung:

Wör-ter-buch, Plural: Wör-ter-bü-cher

Aussprache:

IPA: [ˈvœʁtə, buːχ], Plural: [ˈvœʁtə, byːçɐ]

Hörbeispiele:  Wörterbuch

(österreichisch) [\(Info\)](#), Plural:  Wörterbücher

(österreichisch) [\(Info\)](#)

Bedeutungen:

[1] Nachschlagewerk für die Schreibweise, Bedeutung, Grammatik, Geschichte und/oder Übersetzung von Wörtern

Abkürzungen:

[1] **Wb.**, **Wtb.**

Herkunft:

Verdeutschung von *dictionarium* bzw. *lexicon* im 17. Jahrhundert durch [Comenius](#)^[1], ursprünglich holländisch: *wordboek*.

Synonyme:

[1] *manchmal auch*: [Diktionär](#), [Enzyklopädie](#), [Lexikon](#), [Thesaurus](#)

	Singular	Plural
Nominativ	das Wörterbuch	die Wörterbücher
Genitiv	des Wörterbuchs des Wörterbuches	der Wörterbücher
Dativ	dem Wörterbuch dem Wörterbuche	den Wörterbüchern
Akkusativ	das Wörterbuch	die Wörterbücher



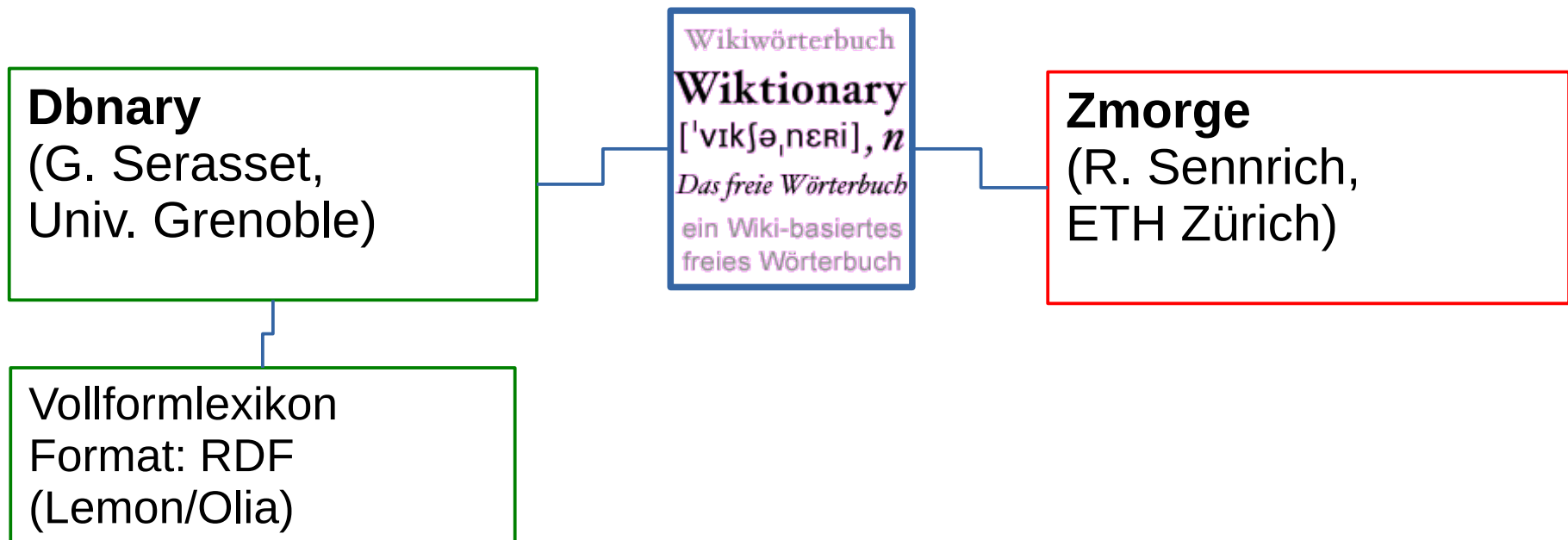
[1] Lateinisches Wörterbuch 

- Morphologische Information

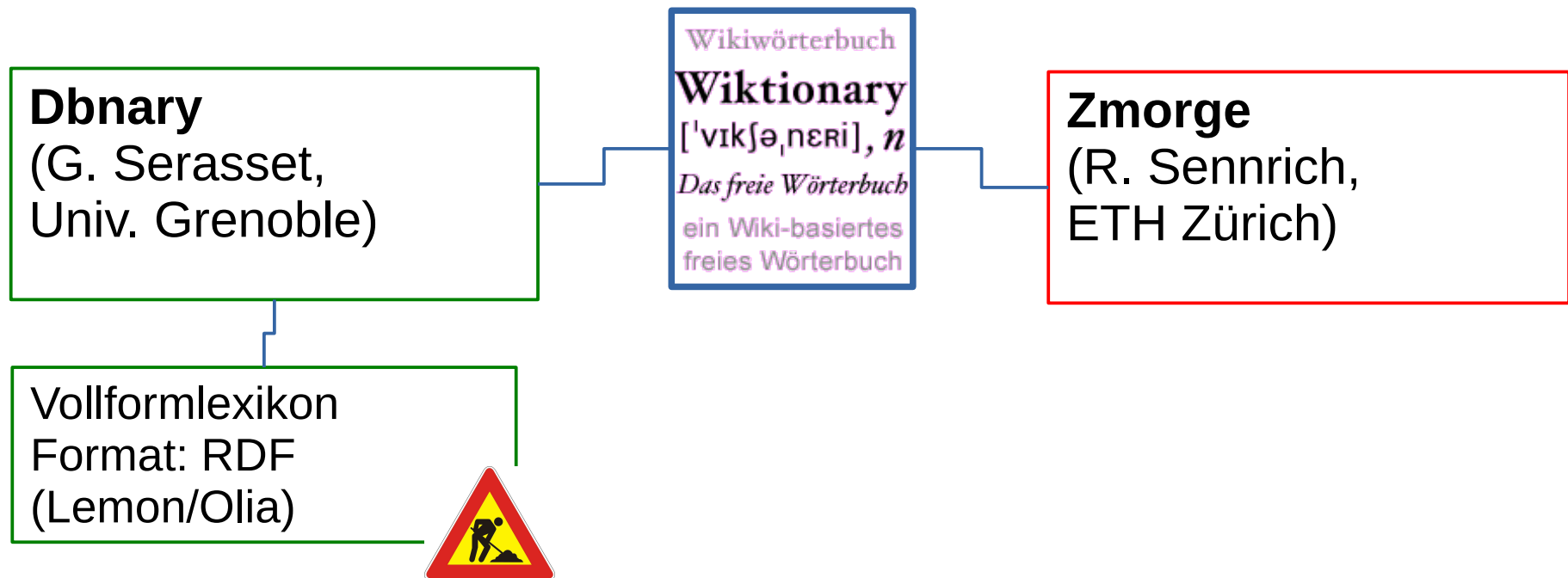
Wiktionary: Extraktion von Informationen aus Flexionstabellen



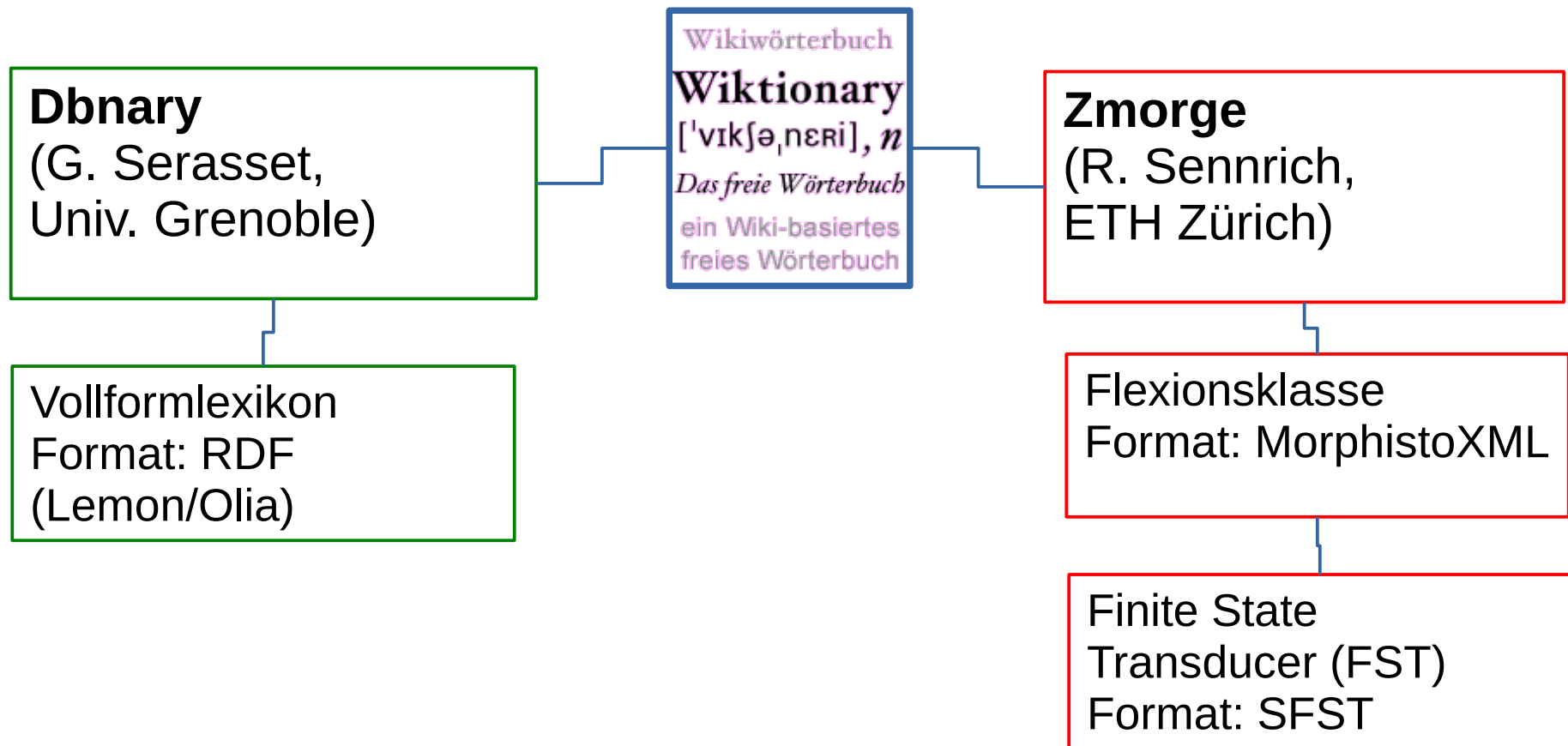
Wiktionary: Extraktion von Informationen aus Flexionstabellen



Wiktionary: Extraktion von Informationen aus Flexionstabellen



Wiktionary: Extraktion von Informationen aus Flexionstabellen



Dbnary (Lemon/RDF) vs. Zmorge (MorphistoXML)

Wiktionary: Flexionstabelle für „schön“

Positiv								
Starke Deklination								
	Singular						Plural	
	Maskulinum		Femininum		Neutrum		—	
	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform
Nominativ	—	schöner	—	schöne	—	schönes	—	schöne
Genitiv	—	schönen	—	schöner	—	schönen	—	schöner
Dativ	—	schönem	—	schöner	—	schönem	—	schönen
Akkusativ	—	schönen	—	schöne	—	schönes	—	schöne
Schwache Deklination								
	Singular						Plural	
	Maskulinum		Femininum		Neutrum		—	
	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform
Nominativ	der	schöne	die	schöne	das	schöne	die	schönen
Genitiv	des	schönen	der	schönen	des	schönen	der	schönen
Dativ	dem	schönen	der	schönen	dem	schönen	den	schönen
Akkusativ	den	schönen	die	schöne	das	schöne	die	schönen

und weitere 10 Tabellen ...

Dbnary (Lemon/RDF) vs. Zmorge (MorphistoXML)

Wiktionary: Flexionstabelle für „schön“

Positiv								
Starke Deklination								
	Singular						Plural	
	Maskulinum		Femininum		Neutrum		—	
	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform
Nominativ	—	schöner	—	schöne	—	schönes	—	schöne
Genitiv	—	schönen	—	schöner	—	schönen	—	schöner
Dativ	—	schönem	—	schöner	—	schönem	—	schönen
Akkusativ	—	schönen	—	schöne	—	schönes	—	schöne
Schwache Deklination								
	Singular						Plural	
	Maskulinum		Femininum		Neutrum		—	
	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform
Nominativ	der	schöne	die	schöne	das	schöne	die	schönen
Genitiv	des	schönen	der	schönen	des	schönen	der	schönen
Dativ	dem	schönen	der	schönen	dem	schönen	den	schönen
Akkusativ	den	schönen	die	schöne	das	schöne	die	schönen

und weitere 10 Tabellen ...

DBnary

```

deu:schön__Adjektiv__1
  lemon:otherForm [ dbnary:note
    Prädikativ"^^<http://www.w3.org/2001/XMLSchema#string> ;
    olia:hasDegree olia:Superlative ;
    olia:hasGender olia:Masculine ;
    olia:hasNumber olia:Singular ;
    lemon:writtenRep "er ist am schönsten"@de
  ] ;
  lemon:otherForm [ olia:hasDegree olia:Comparative ;
    olia:hasGender olia:Masculine ;
    olia:hasInflectionType olia:StrongInflection ;
    olia:hasNumber olia:Singular ;
    lemon:writtenRep "schönerem"@de
  ] ;
  lemon:otherForm [ olia:hasDegree olia:Comparative ;
    olia:hasGender olia:Masculine ;
    olia:hasInflectionType olia:StrongInflection ;
    olia:hasNumber olia:Singular ;
    lemon:writtenRep "schöneren"@de
  ] ;
  ...

```

Dbnary (Lemon/RDF) vs. Zmorph (MorphistoXML)

Wiktionary: Flexionstabelle für „schön“

Positiv								
Starke Deklination								
	Singular						Plural	
	Maskulinum		Femininum		Neutrum		—	
	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform
Nominativ	—	schöner	—	schöne	—	schönes	—	schöne
Genitiv	—	schönen	—	schöner	—	schönen	—	schöner
Dativ	—	schönem	—	schöner	—	schönem	—	schönen
Akkusativ	—	schönen	—	schöne	—	schönes	—	schöne
Schwache Deklination								
	Singular						Plural	
	Maskulinum		Femininum		Neutrum		—	
	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform
Nominativ	der	schöne	die	schöne	das	schöne	die	schönen
Genitiv	des	schönen	der	schönen	des	schönen	der	schönen
Dativ	dem	schönen	der	schönen	dem	schönen	den	schönen
Akkusativ	den	schönen	die	schöne	das	schöne	die	schönen

und weitere 10 Tabellen ...

Dbnary (Lemon/RDF) vs. Zmorge (MorphistoXML)

Wiktionary: Flexionstabelle für „schön“

Positiv								
Starke Deklination								
	Singular						Plural	
	Maskulinum		Femininum		Neutrum		—	
	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform
Nominativ	—	schöner	—	schöne	—	schönes	—	schöne
Genitiv	—	schönen	—	schöner	—	schönen	—	schöner
Dativ	—	schönem	—	schöner	—	schönem	—	schönen
Akkusativ	—	schönen	—	schöne	—	schönes	—	schöne
Schwache Deklination								
	Singular						Plural	
	Maskulinum		Femininum		Neutrum		—	
	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform	Artikel	Wortform
Nominativ	der	schöne	die	schöne	das	schöne	die	schönen
Genitiv	des	schönen	der	schönen	des	schönen	der	schönen
Dativ	dem	schönen	der	schönen	dem	schönen	den	schönen
Akkusativ	den	schönen	die	schöne	das	schöne	die	schönen


Zmorge / MorphistoXML

```

<BaseStem>
  <Lemma>schön</Lemma>
  <Stem>schön</Stem>
  <Pos>ADJ</Pos>
  <Origin>nativ</Origin>
  <InfClass>Adj+</InfClass>
</BaseStem>
  
```

und weitere 10 Tabellen ...

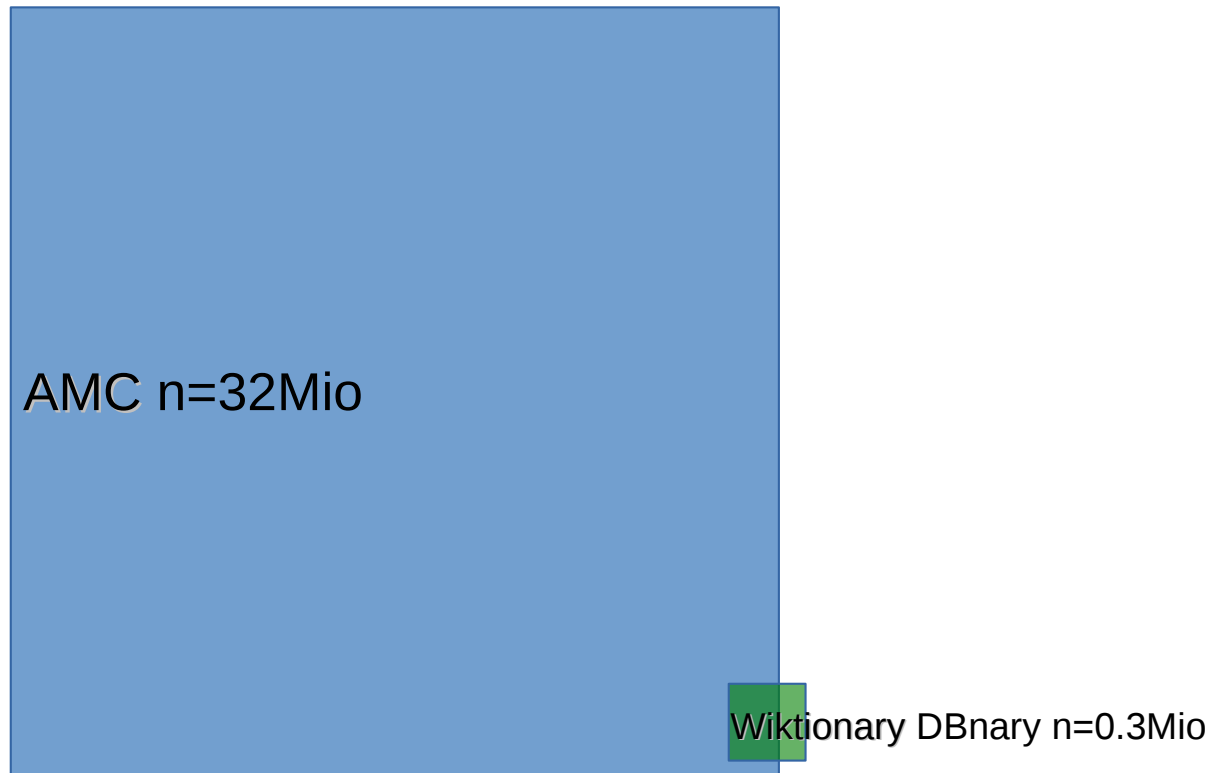
Austrian Media Corpus (AMC)

- Quelle: Austria Press Agentur (APA) 
- (österreichische) OTS Meldungen, Zeitungen und Magazine
- 1989-2013
- 33 Mio Dokumente
- 6.1 Mrd Wörter
- 32 Mio Wortformen
- POS annotiert (Tree Tagger, RF Tagger)
- 450.000 erkannte Lemmata

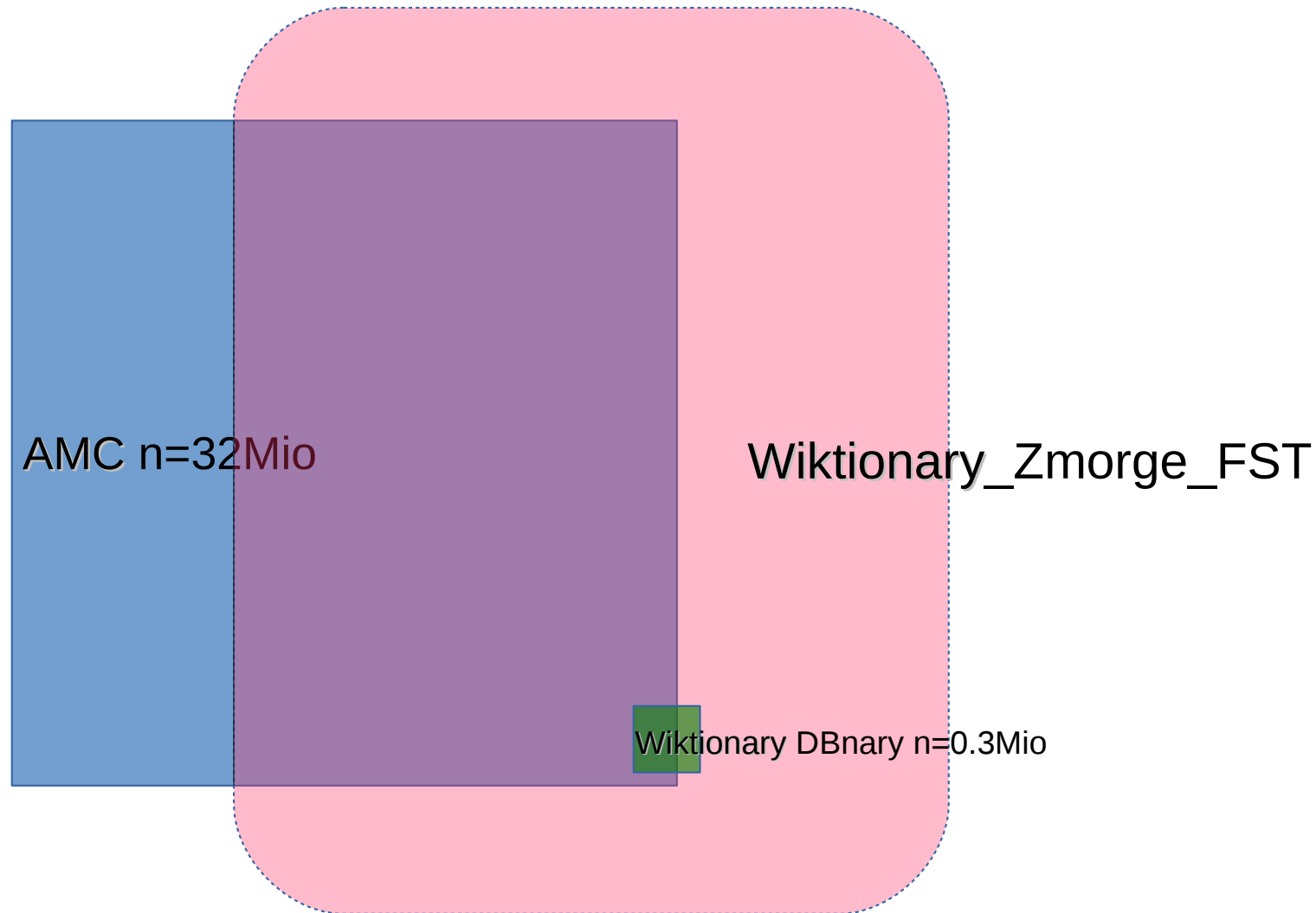
Wiktionary & AMC



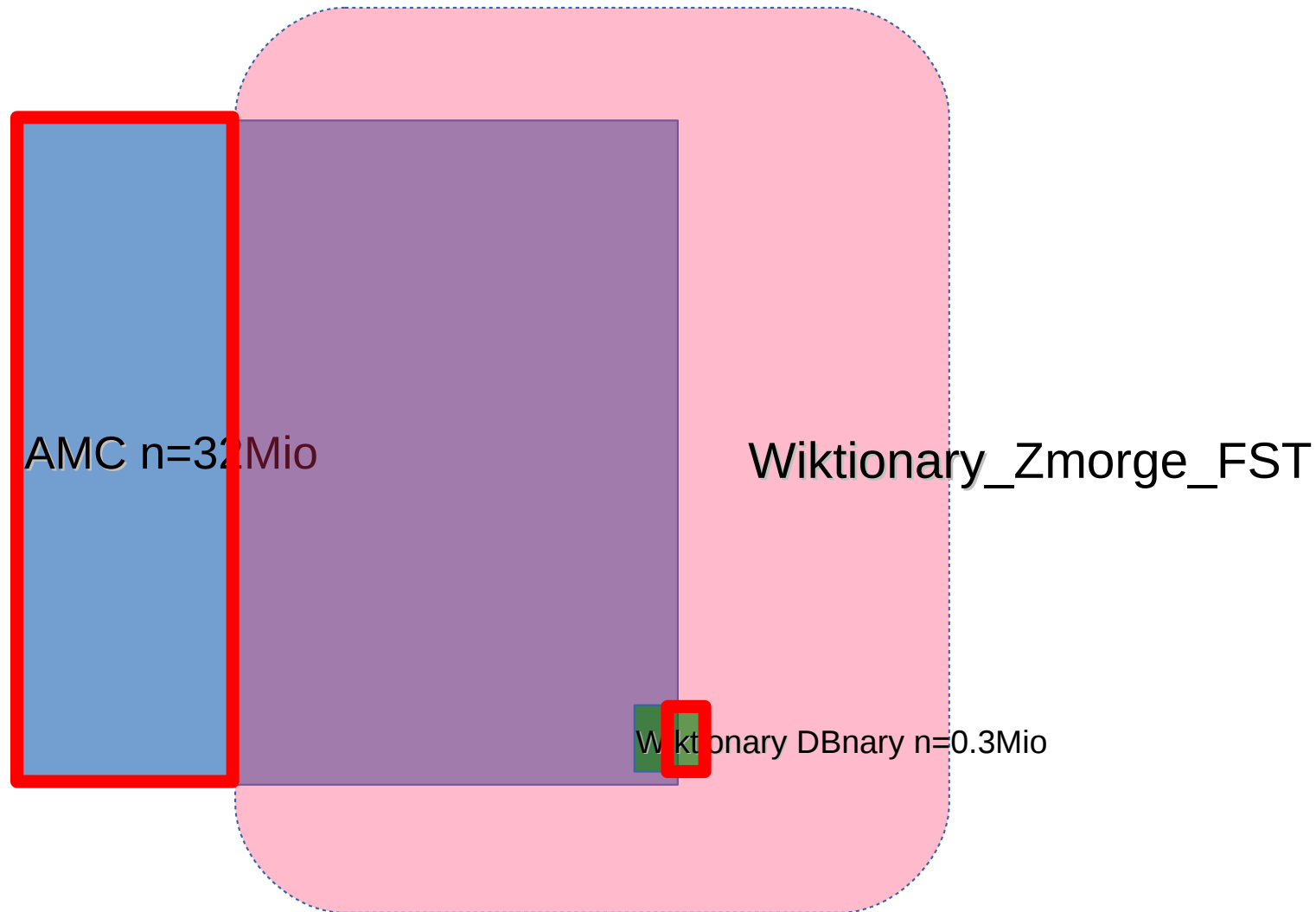
Wiktionary & AMC



Wiktionary & AMC



Wiktionary & AMC: Lücken



Lakunen in Zmorge: Beispiele

- APA
- SPÖ
- Heinz
- Star
- trainieren, stürmen,
punkten, übermitteln
- erstaunlich,
sensationell, riskant

Lakunen in Zmorge: Beispiele

- APA
- SPÖ
- Heinz
- Star
- trainieren, stürmen, punkten, übermitteln
- erstaunlich, sensationell, riskant

Heinz (Deutsch) [Bearbeiten]

Substantiv, Vorname [Bearbeiten]

Hannes (Deutsch) [B

Substantiv, *m*, Vorname

Lakunen in Zmorge: Beispiele

- APA
- SPÖ
- Heinz
- Star
- trainieren, stürmen, punkten, übermitteln
- erstaunlich, sensationell, riskant

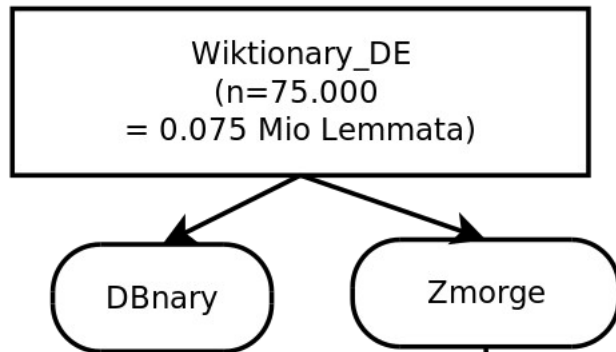
	Singular	Plural
Nominativ	der Star	die Stare die Stars
Genitiv	des Staren des Stares des Stars	der Stare der Stars
Dativ	dem Star dem Stare	den Staren den Stars
Akkusativ	den Star	die Stare die Stars



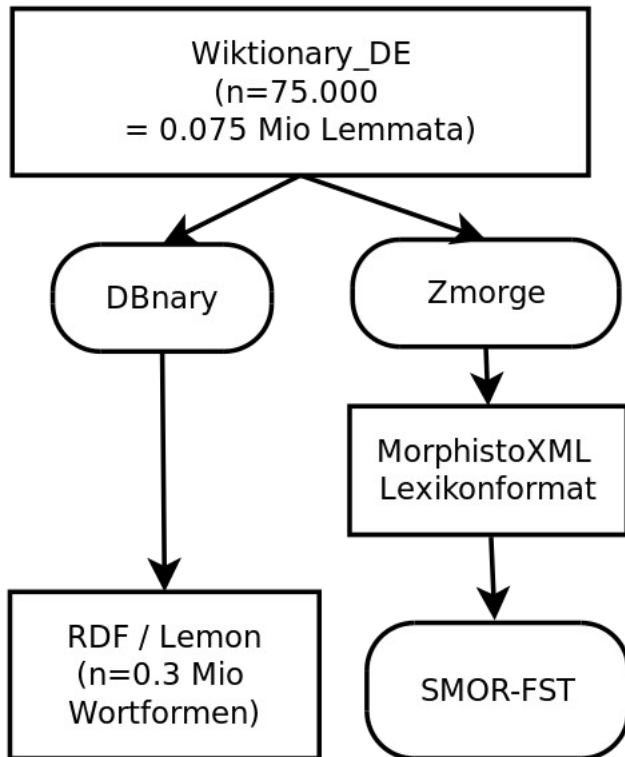
Aktualität von Wiktionary

- Aktualisierungen in jüngster Duden-Auflage:
 - twittern
 - fremschämen
 - Abwrackprämie
 - Zwergplanet
- sämtliche auch in Wiktionary
- in AMC:
 - twittern, zugetwittert, vertwittert, betwittert, ausgetwittert, ...

Baffle-Generierung: Workflow



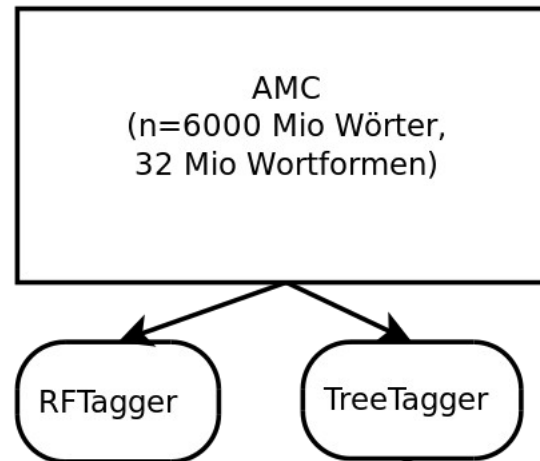
Baffle-Generierung: Workflow



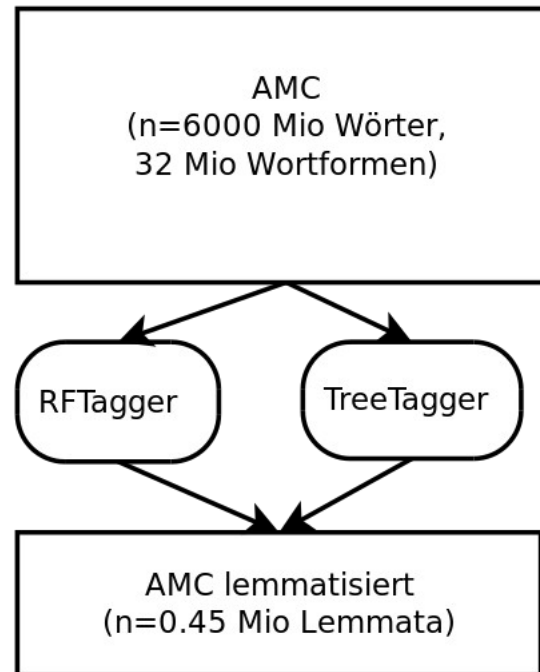
Baffle-Generierung: Workflow

AMC
(n=6000 Mio Wörter,
32 Mio Wortformen)

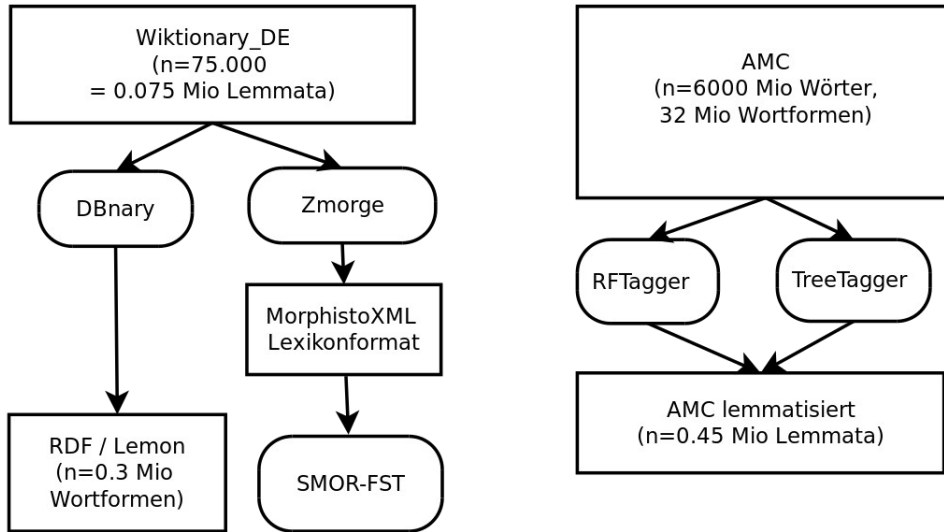
Baffle-Generierung: Workflow



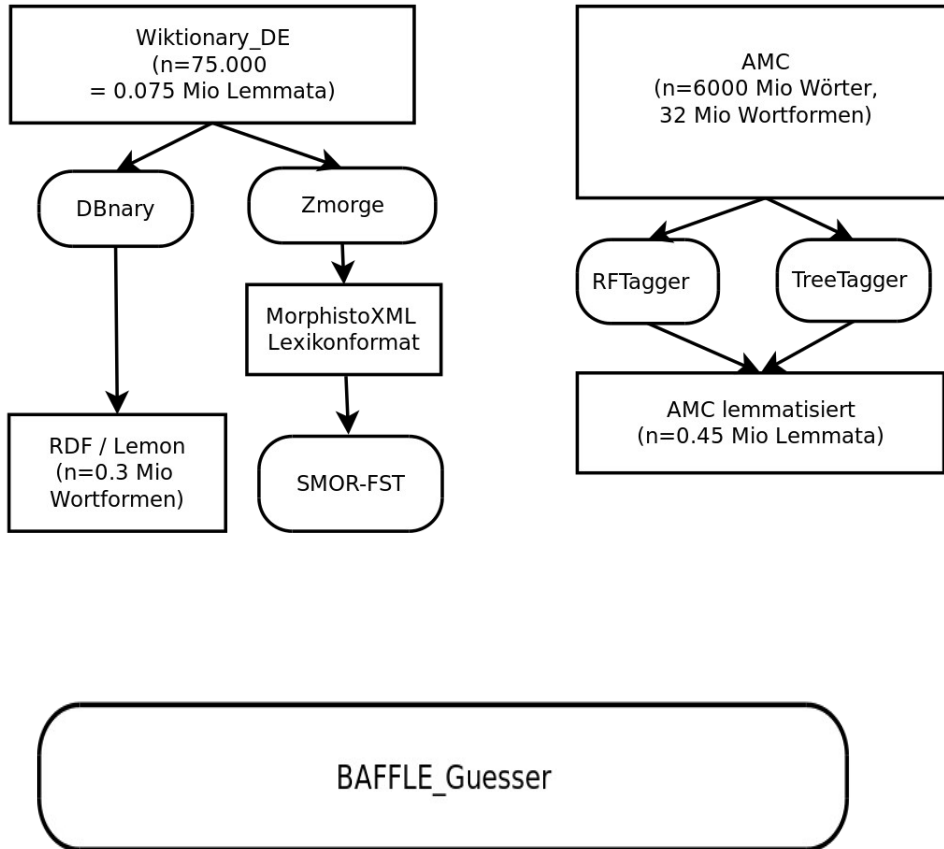
Baffle-Generierung: Workflow



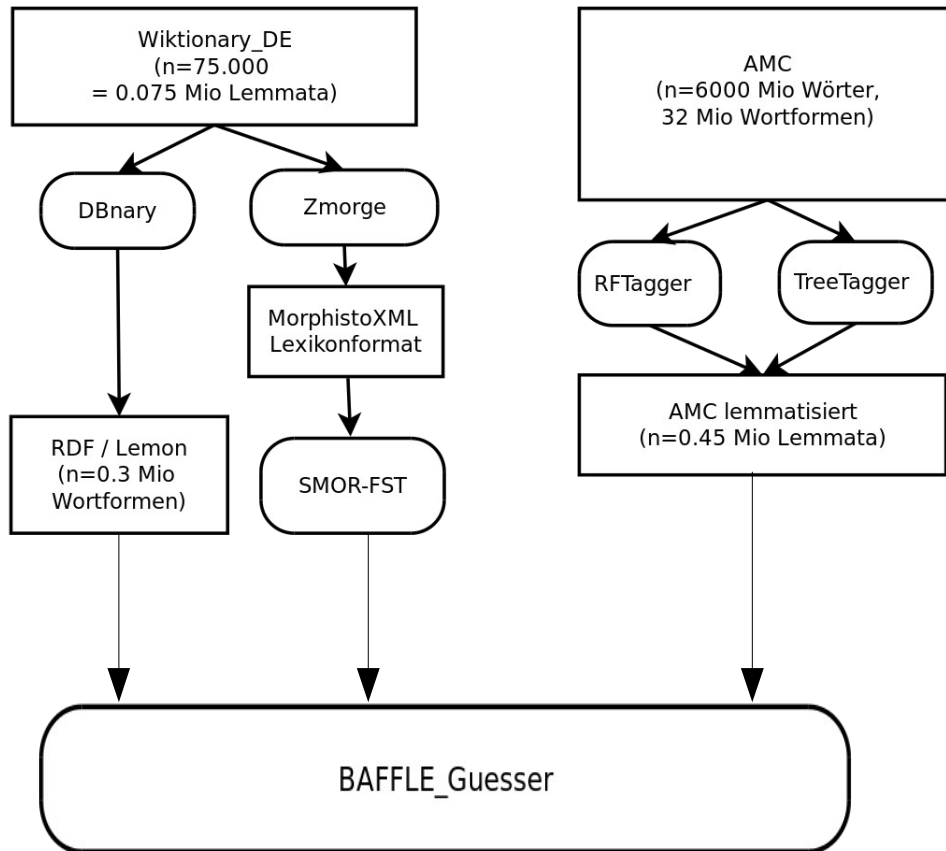
Baffle-Generierung: Workflow



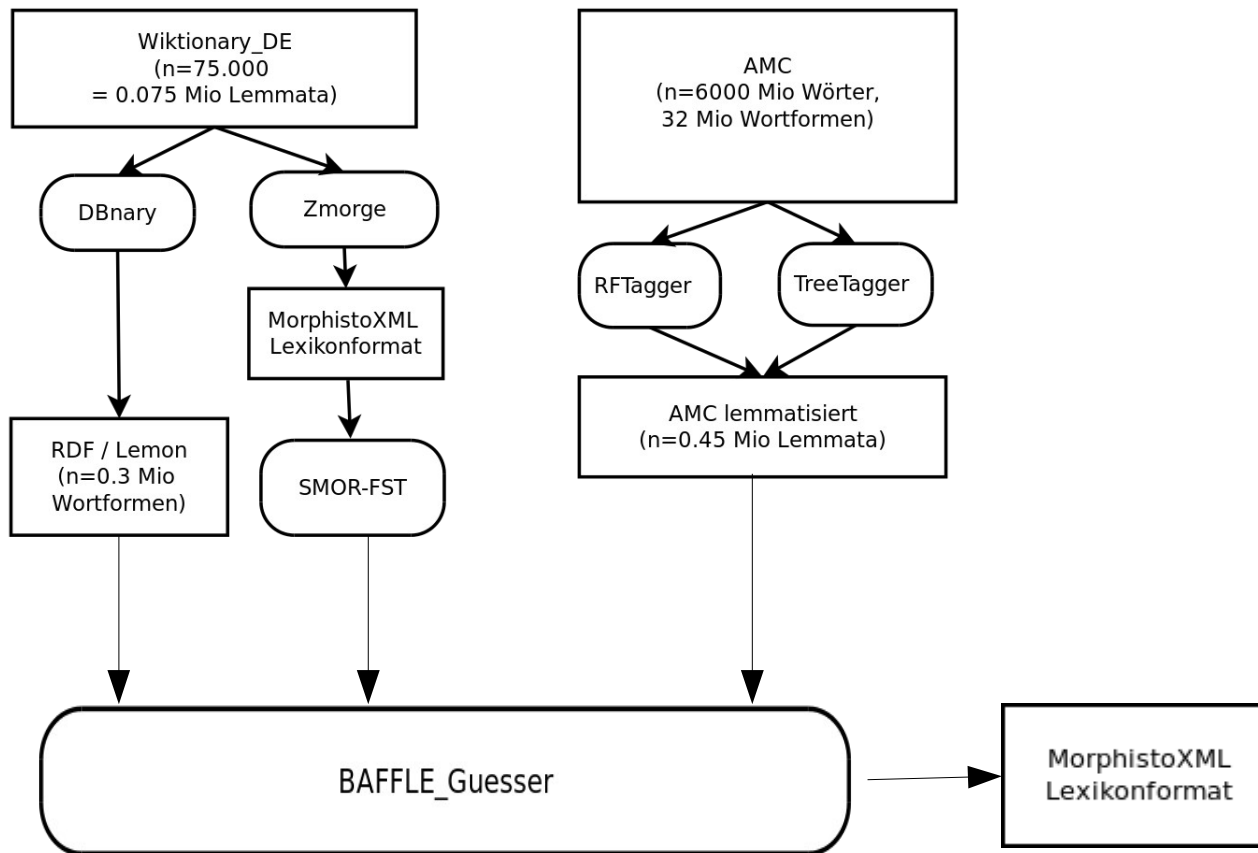
Baffle-Generierung: Workflow



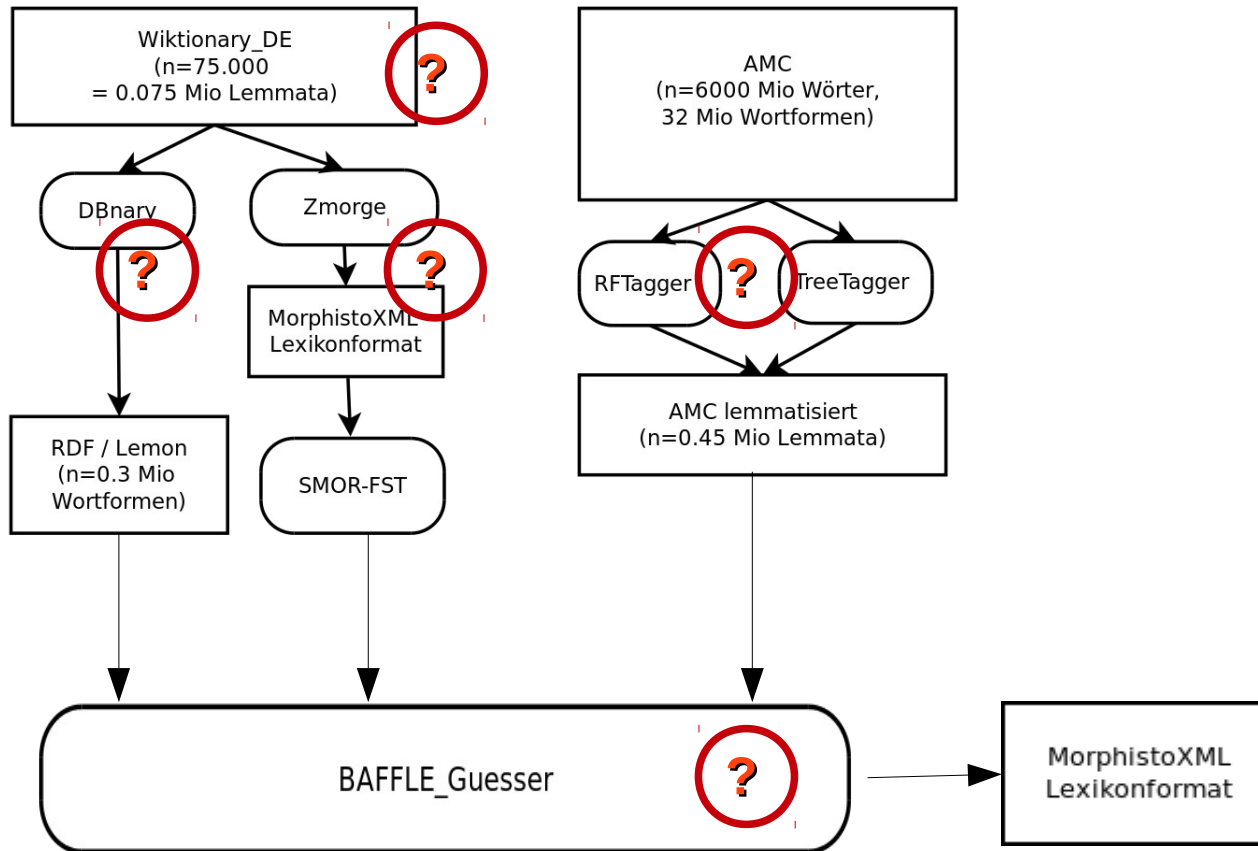
Baffle-Generierung: Workflow



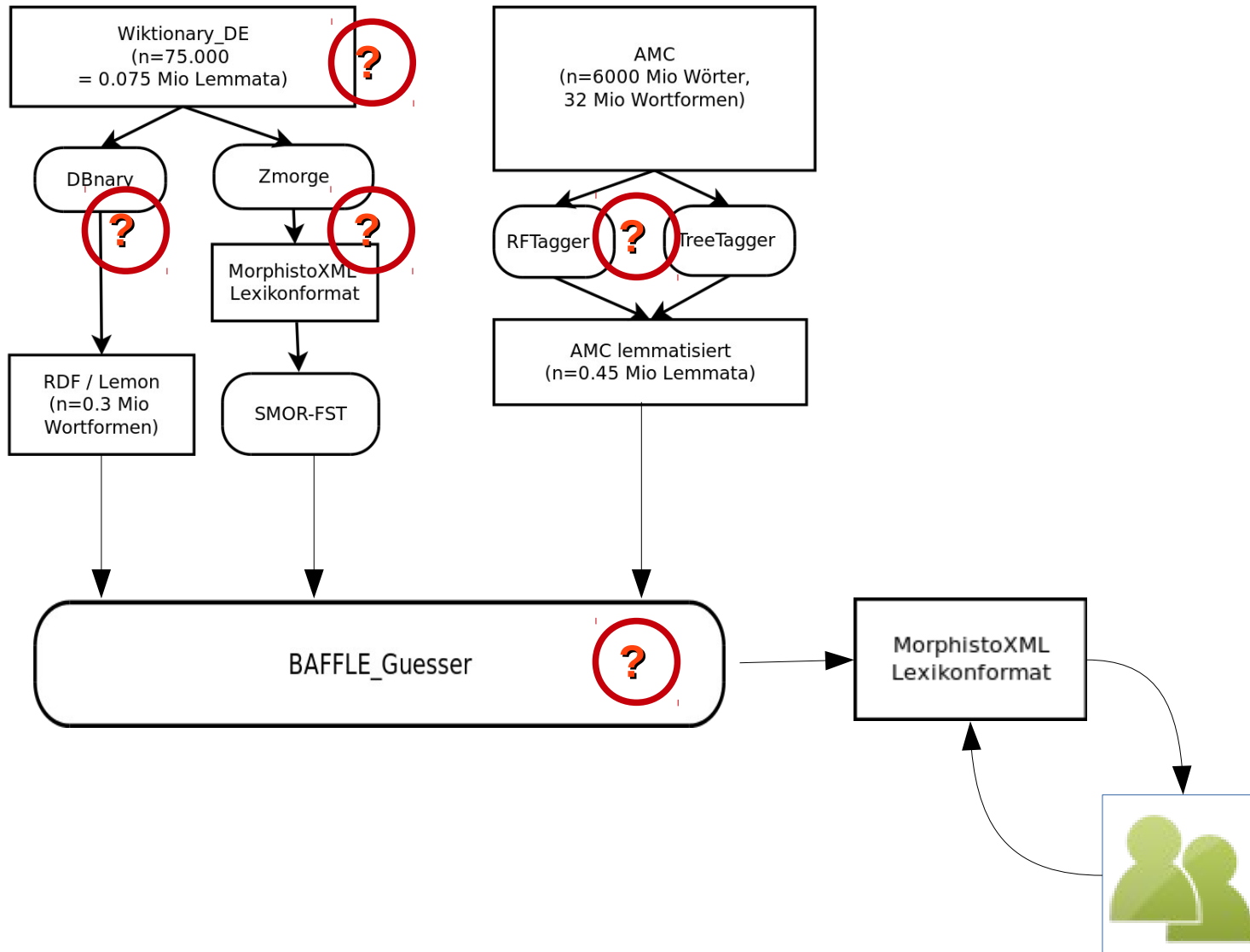
Baffle-Generierung: Workflow



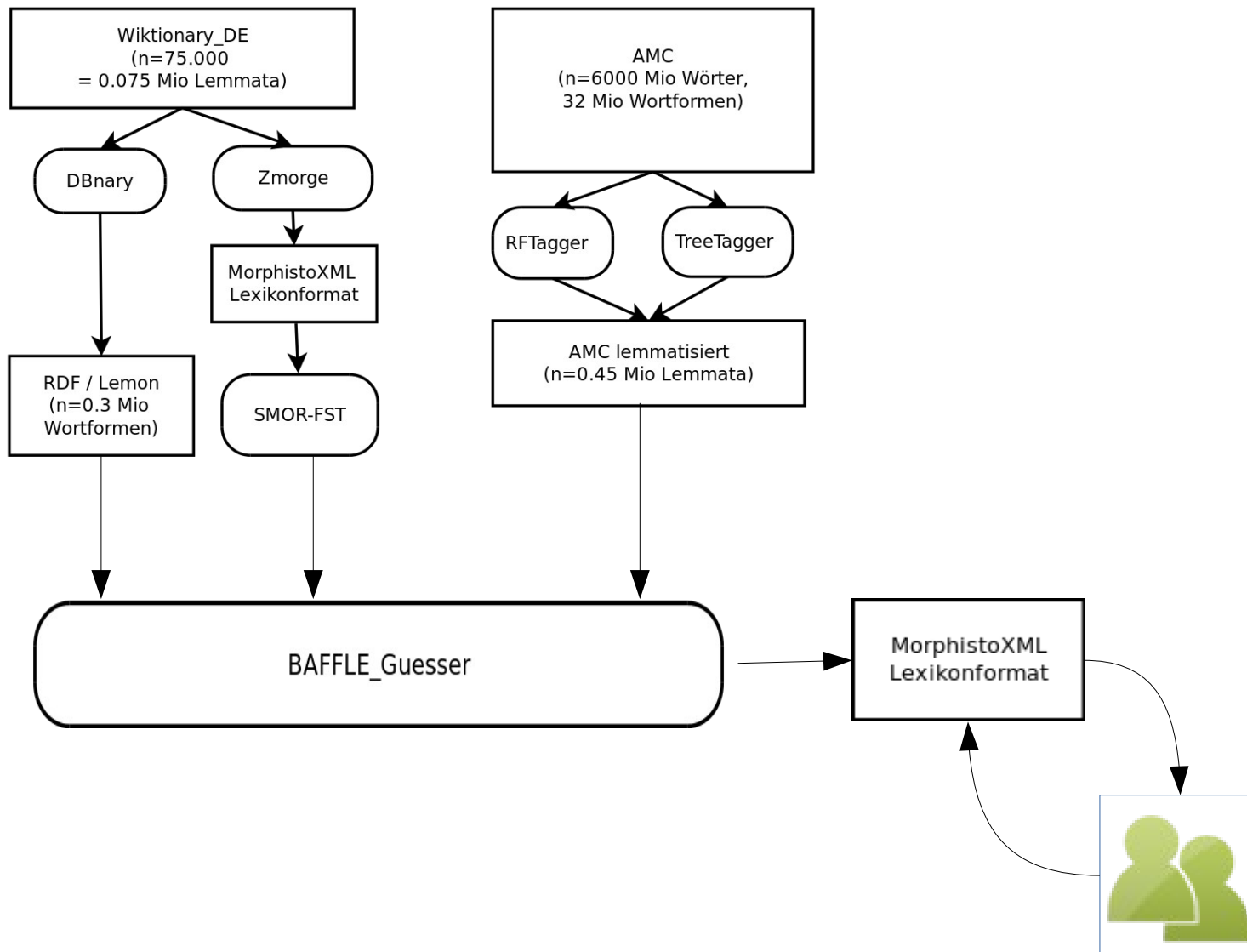
Baffle-Generierung: Fehlerquellen



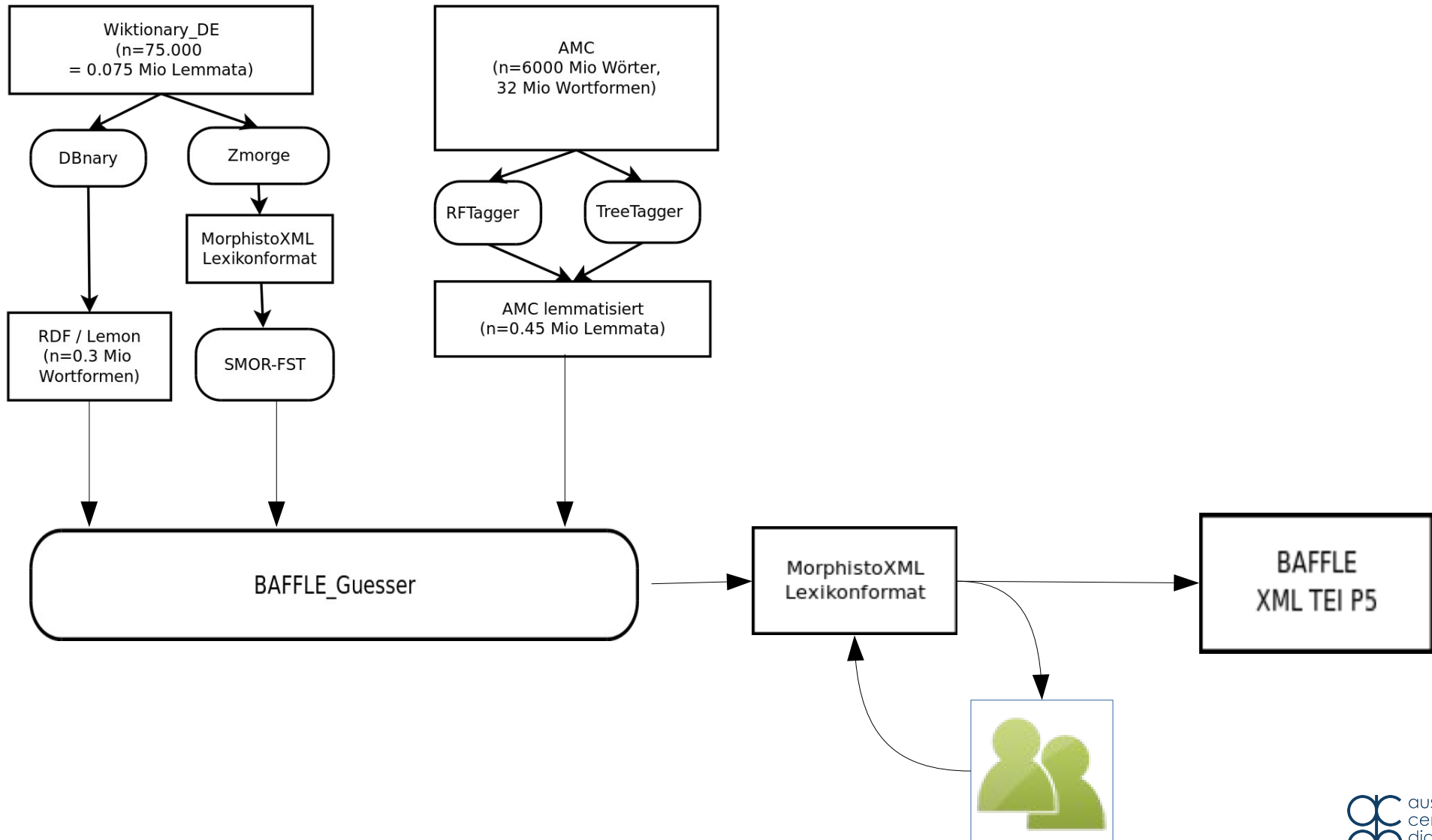
Baffle-Generierung: Fehlerkorrektur



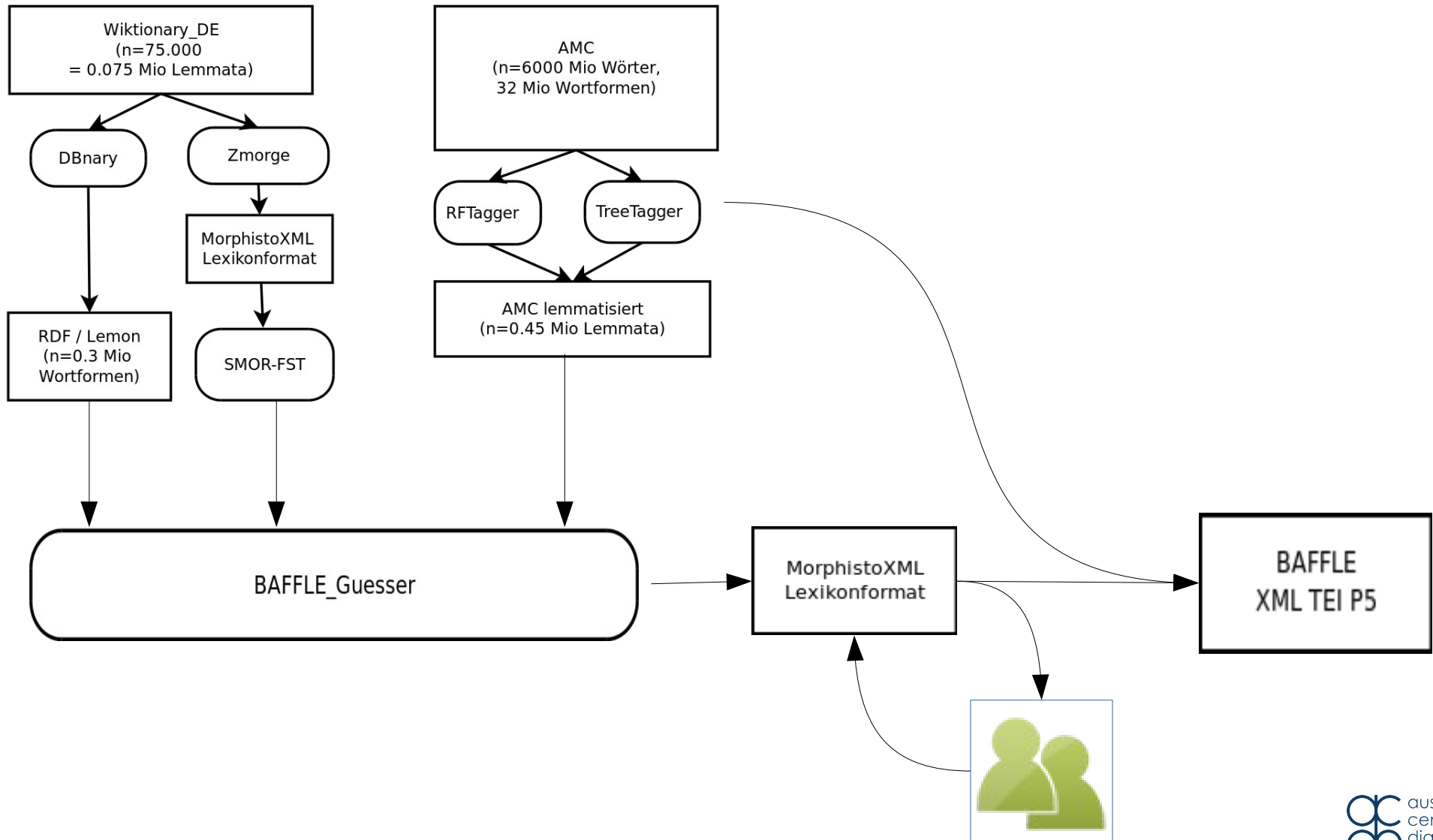
Baffle-Generierung: Fehlerkorrektur



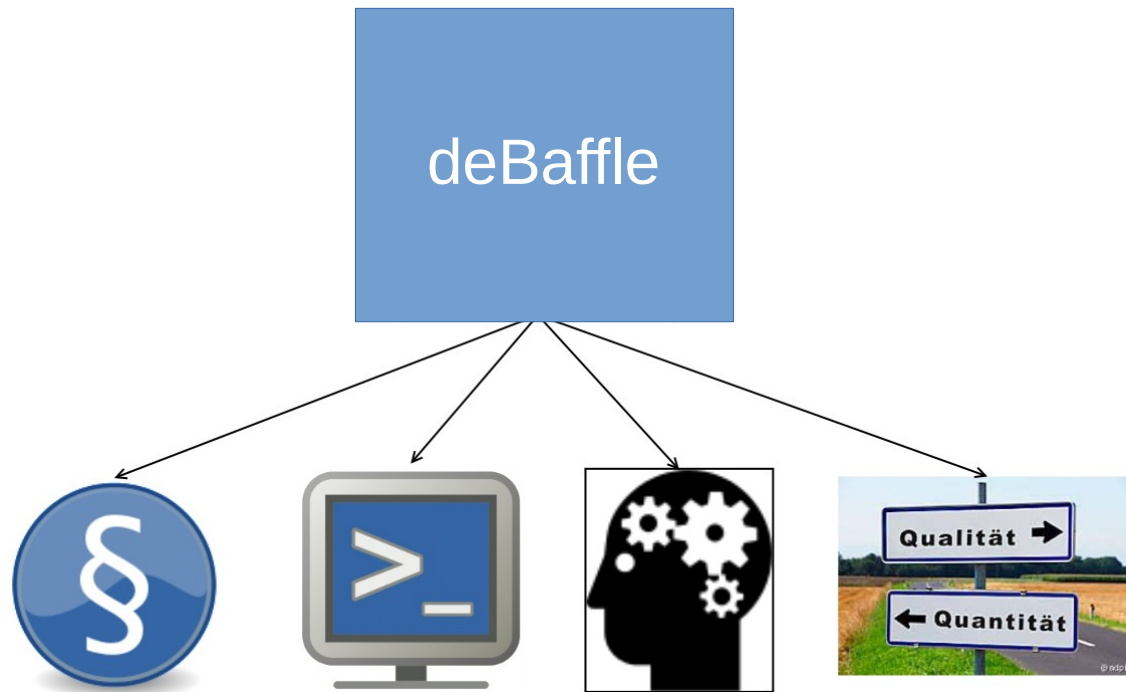
Baffle-Generierung: TEI XML Output



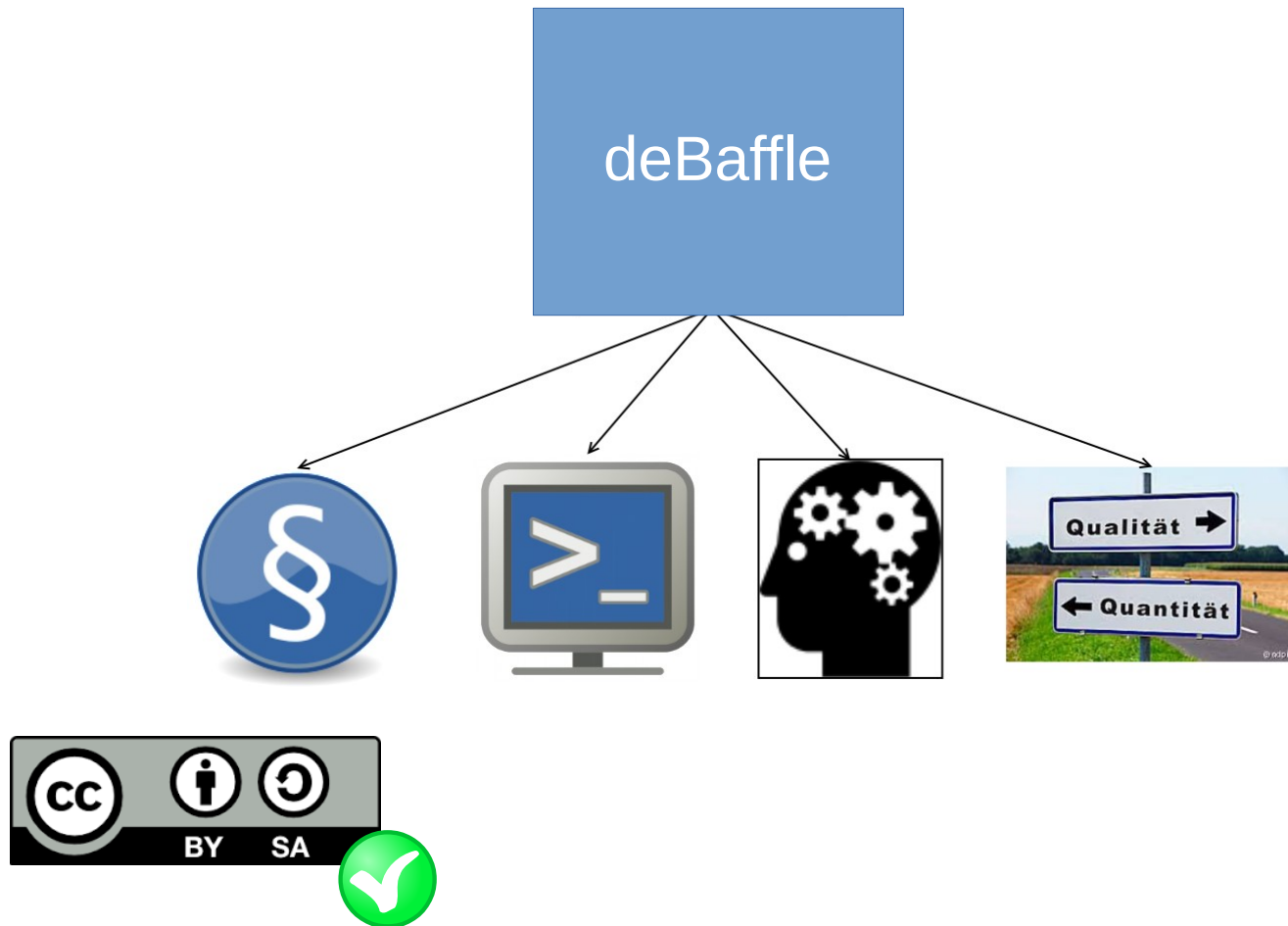
Baffle-Generierung: Frequenzdaten



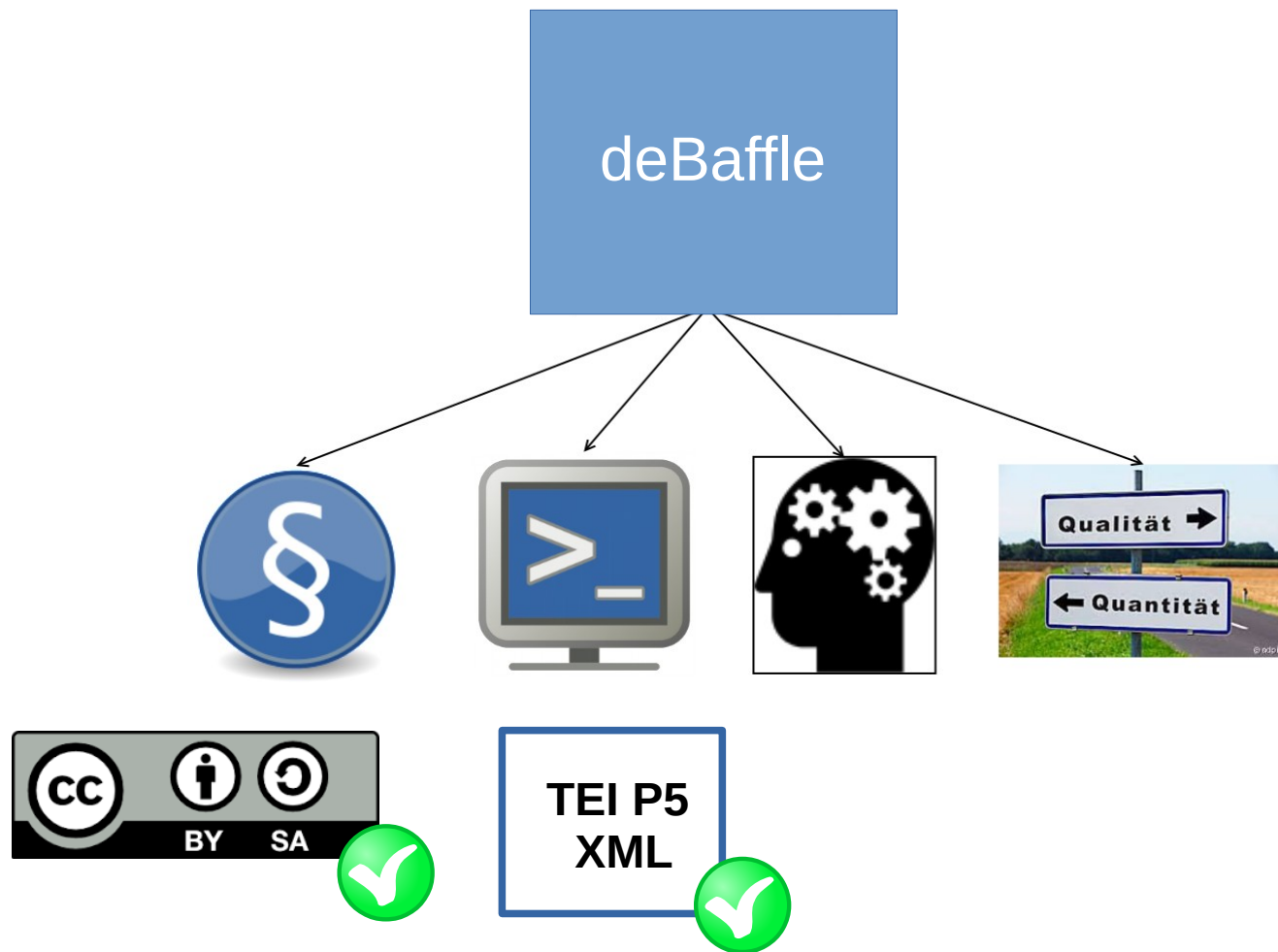
Zusammenfassung



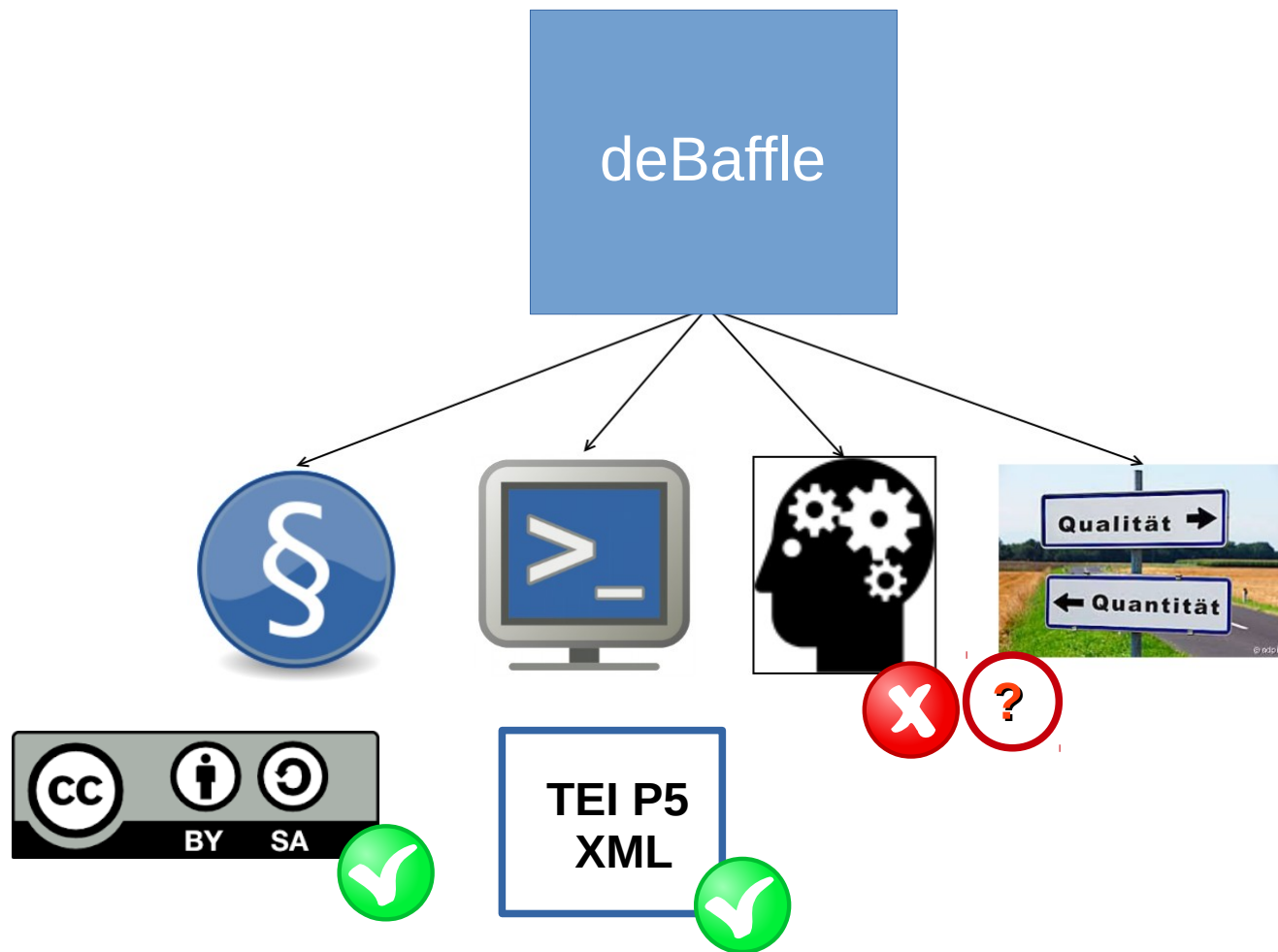
Zusammenfassung



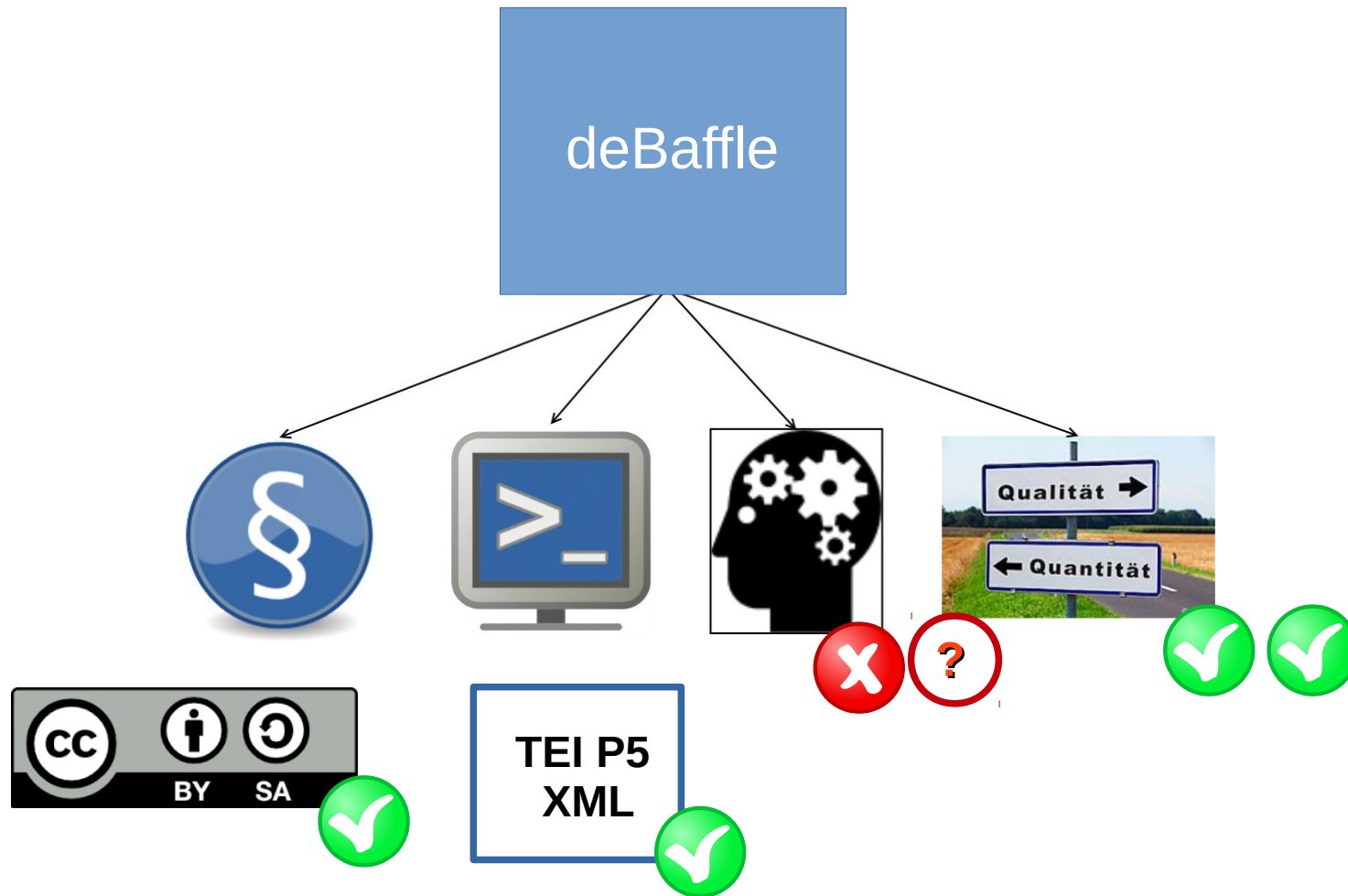
Zusammenfassung



Zusammenfassung



Zusammenfassung



Danke für Ihre Aufmerksamkeit