

Literaturarbeit XML-Schema-Evolution: Kategorisierung und Bewertung

Hannes Grunert



Einführendes Beispiel

Angestrebter Abschluss	Studiengang	Fachsemester
Bachelor	Informatik	9
Master	Informatik	1

```
<studium>  
  <studienfach>  
    <abschluss>Bachelor</abschluss>  
    <studiengang>Informatik</studiengang>  
    <semester>9</semester>  
  </studienfach>  
  <studienfach>  
    <abschluss>Master</abschluss>  
    <studiengang>Informatik</studiengang>  
    <semester>1</semester>  
  </studienfach>  
</studium>
```

Einführendes Beispiel – XML-Schema (Ausschnitt)

```
<xsd: element name="studium">  
  <xsd:complexType>  
    <xsd: sequence>  
      <xsd:element ref="studienfach " minOccurs="1" maxOccurs="unbounded"/>  
    </xsd:sequence>  
  </xsd:complexType>  
</xsd: element>
```

 Evolution

```
<xsd: element name="studium">  
  <xsd:complexType>  
    <xsd: sequence>  
      <xsd:element ref="studienfach " minOccurs="1" maxOccurs="1"/>  
    </xsd:sequence>  
  </xsd:complexType>  
</xsd: element>
```

Einführendes Beispiel

Angestrebter Abschluss	Studiengang	Fachsemester
Bachelor	Informatik	9

```
<studium>  
  <studienfach>  
    <abschluss>Bachelor</abschluss>  
    <studiengang>Informatik</studiengang>  
    <semester>9</semester>  
  </studienfach>  
  <!-- Eintrag durch Evolution entfernt -->  
</studium>
```

Inhalt

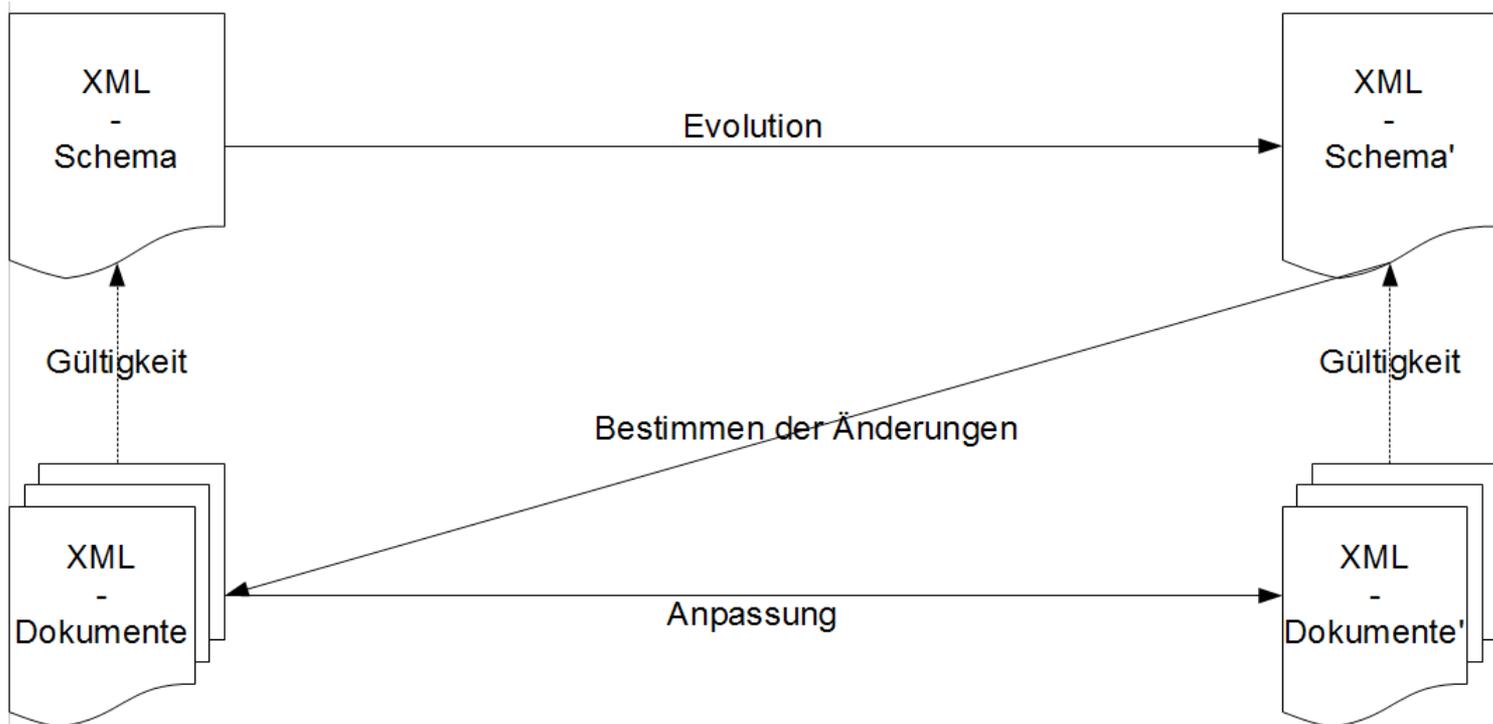
I. Literatararbeit (Kategorisierung)

- i. XML-Schema-Evolution
- ii. Kategorisierungsbegriff
- iii. Kriterien
- iv. Beispiel
- v. Kategorisierungen
- vi. Vergleich

II. Ausblick Bachelorarbeit (Bewertung)

XML-Schema-Evolution

- Neue Version eines XML-Schemas
- Überprüfen der Gültigkeit von Instanzen
- Ggf. Anpassung der Instanzen



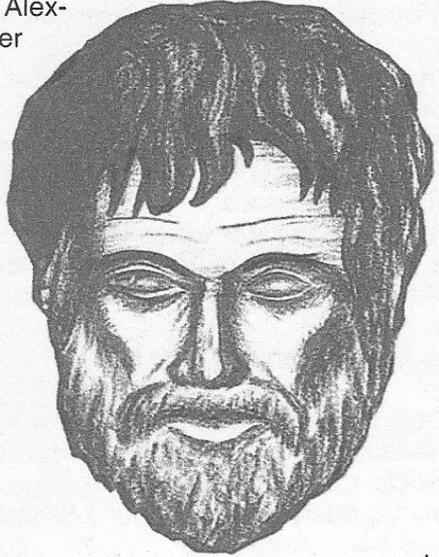
Kategorisierung

- Ursprung: Antike
- Begründer: Aristoteles
- philosophische Grundfrage:
Was ist das Wesen selbst?
- Objekte mit 10 Eigenschaften

Substanz: Aristoteles

Qualität: Philosoph Quantität: 171 cm

Relation: Lehrer
des Alex- Wo: Athen
ander



Wann: Lage:
morgens stehend

Haben: gelassen Tätigkeit: lehren

Leiden: wird verbannt

Kategorien des Aristoteles (aus [DTV])



Kategorisierungskriterien Schema-Evolution

- Was wird geändert? (SimpleType, ComplexType, Attribute)
- Wie wird es verändert? (Löschen, Hinzufügen, Update)
- **Werden Instanzen verändert?**
- **Änderung der Informationskapazität**

Beispiel – Schema (Ausschnitt)

```
<xsd:element name="Person">  
  <xsd:complexType>  
    <xsd:sequence>  
      <xsd:element ref="Name"/>  
      <xsd:element ref="Gesinnung"/>  
    </xsd:sequence>  
  </xsd:complexType>  
</xsd:element>
```



```
<xsd:element name="Person">  
  <xsd:complexType>  
    <xsd:attribute name="Name" type="xsd:string"/>  
    <xsd:sequence>  
      <xsd:element ref="Gesinnung"/>  
    </xsd:sequence>  
  </xsd:complexType>  
</xsd:element>
```

Beispiel - Instanz

```
<Person>  
  <Name>"Thomas"</Name>  
  <Gesinnung>rechtschaffen neutral</Gesinnung>  
</Person>
```

 Transformation

```
<Person name="Thomas">  
  <Gesinnung>rechtschaffen neutral</Gesinnung>  
</Person>
```

Beispiel - Kategorisierung

- Was wird geändert?
 - Complex Type
- Wie wird es verändert?
 - Umwandlung eines Elementes in ein Attribut
- **Werden Instanzen verändert?**
 - **Ja**
- **Änderung der Informationskapazität**
 - **erhaltend**

Kategorisierung nach Su, Kramer et al. ([HS01])

- DTD Change Primitives
 - Changes to the Document Definition
 - Create element with name u
 - Destroy element with name u
 - Rename element from name u to u'
 - Changes to an Element Type Definition
 - Add element E at position pos to parent P with quantifier q and default value d
 - Remove sub-element at position pos in parent E
 - Change quantifier of content particle at position pos in parent E to quantifier q with default value d
 - Group sub-elements from position $start$ to position end in parent E into a list group
 - Flatten group at position pos in element E to a list of sub-elements
 - Change quantifier of group at position pos in element E to q
 - Add attribute with name u to element E with type t , default type d , and default value v
 - Destroy attribute with name u from element E
 - Changes to an Attribute Type Definition
 - Change element E 's attribute u 's type to t , with default value v
 - Change element E 's attribute u 's default value to v
 - Change element E 's attribute u 's fixed value to v
- Complex Changes

Kategorisierung nach Tan und Goh ([MT04])

Migratory Changes

- Morphing of an element to an attribute.
- Migration of a sub-element from one element to another.
- Modification of an attribute or element.

Structural Changes

- Addition of new elements, sub-elements and attributes.
- Removal of elements, sub-elements and attributes.

Sedentary Changes

- Renaming of elements.
 - Renaming of attributes
 - Change of simple data type.
-

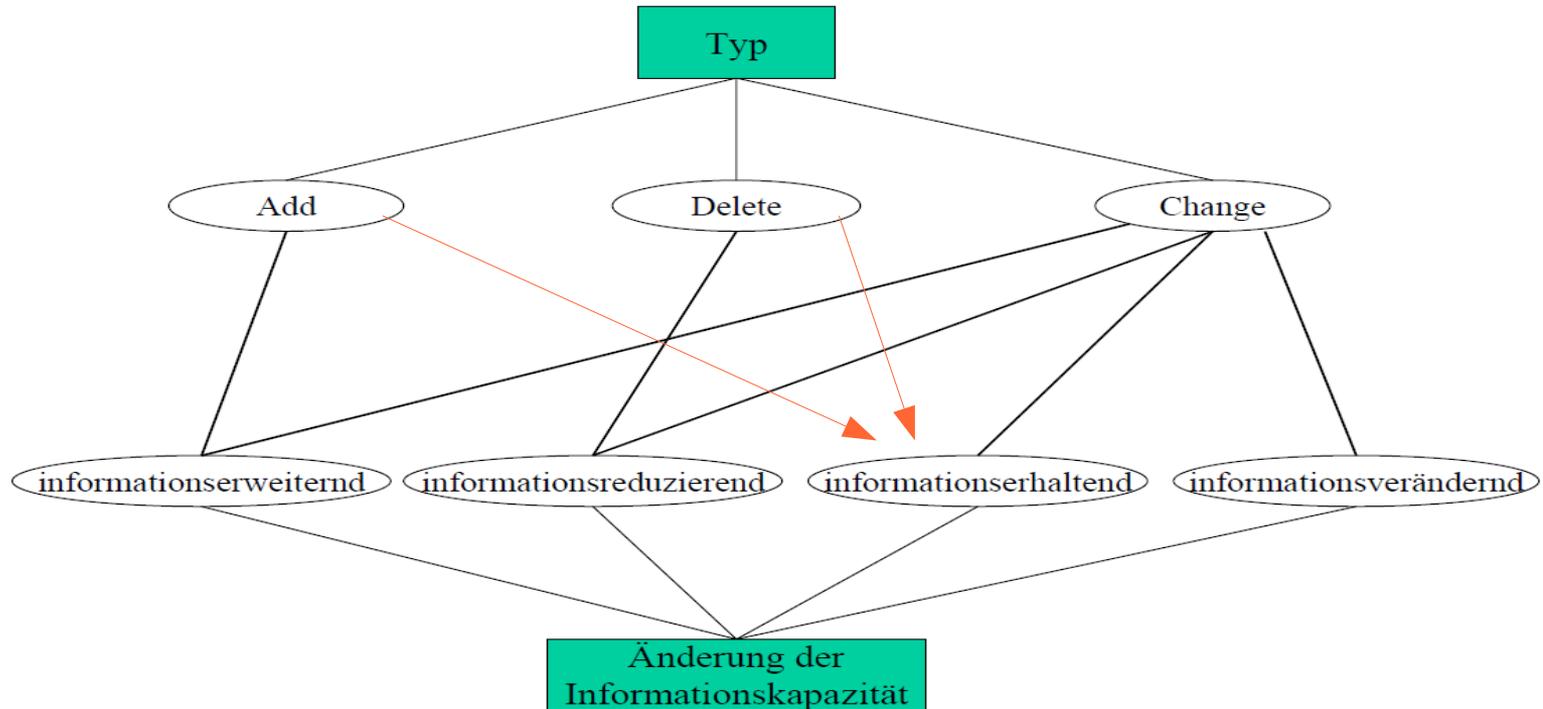
- Art der Änderung in 3 Hauptkategorien
- Unterkategorien mit Typ

Kategorisierung nach Guerrini, Mesiti und Rossi ([GG06])

	Insertion	Modification	Deletion
Simple Type	<i>insert_glob_simple_type*</i> <i>insert_new_member_type*</i>	<i>change_restriction</i> <i>change_base_type</i> <i>rename_type*</i> <i>change_member_type</i> <i>global_to_local*</i> <i>local_to_global*</i>	<i>remove_type*</i> <i>remove_member_type*</i>
Complex Type	<i>insert_glob_complex_type*</i> <i>insert_local_elem</i> <i>insert_ref_elem</i> <i>insert_operator</i>	<i>rename_local_elem</i> <i>rename_global_type*</i> <i>change_type_local_elem</i> <i>change_cardinality</i> <i>change_operator</i> <i>global_to_local*</i> <i>local_to_global*</i>	<i>remove_element</i> <i>remove_operator</i> <i>remove_substructure</i> <i>remove_type*</i>
Element	<i>insert_glob_elem</i>	<i>rename_glob_elem*</i> <i>change_type_glob_elem</i> <i>ref_to_local*</i> <i>local_to_ref*</i>	<i>remove_glob_elem*</i>

- Art der Änderung in 3 Kategorien
- Typ in 3 Kategorien
- Änderungsoperationen mit * haben keine Auswirkungen auf Instanzen

Kategorisierung nach Hartung ([Har07])



- Art der Änderung in 3 Kategorien
- Änderung der Informationskapazität
- Abweichungen von der Original-Abbildung:

Kategorisierung nach Nečaský und Mlýnková ([MN09])

- Structural:
 - Adding – adds a new item
 - Removal – removes a new item
- Sedentary:
 - Extension – adds a new item that does not change structure
 - Renaming – renames an item
 - Renumbering – changes the cardinality of an item
 - Retyping – changes the data type of an item
 - Resetting – changes the value of an item
 - Mapping – maps an item to an item from another level
 - Unmapping – removes a mapping between levels
- Migratory:
 - Moving – moves an item
 - Reordering – changes the order of a set of items
 - Transformation – transforms an item to an item of a different type

Kategorisierung nach Malý ([Mal10])

- addition - new construct was added
- removal - a construct was removed from the model
- migratory - a construct (and its subtree in the XSem-H tree) was moved to another part of the XSem-H tree
- sedentary - an existing construct was adjusted in place, but not moved
- →Aufspaltung der vorherigen structural-Kategorie in addition und removal

Kategorisierung nach Will ([Wil06])

- Umfangreichste Sammlung von Evolutionsschritten
 - Bezieht Beziehungen, Identitätsdefinitionen und Anmerkungen mit ein
- 1. Stufe: Was wird geändert?
- 2. Stufe: Wie wird es geändert
- Kurze Anmerkung, dass auch nach Auswirkungen auf Instanzen kategorisiert werden kann

Vergleich

- Betrachtete Kategorisierungen erfüllen nicht alle Kriterien oder beziehen sich nicht auf XML-Schema als Schemasprache
- Menge der Evolutionsschritte von Christian Will am umfangreichsten

Kategorisierung	Sprache	Vollständigkeit	Komponente	Art	Kapazität	Auswirkung
XEM	DTD	+	+	+	-	+
Andre Zeitz	DTD	+	+	+	+	+
Tan, Goh	XML-Schema	-	+	+	-	-
Guerrini, Mesiti, Rossi	XML-Schema	-	+	+	-	+
Universität Leipzig	XML-Schema	-	-	+	+	(+)
Moro, Malaika, Lim	XML-Schema	-	(+)	+	-	-
Cicchetti, Di Ruscio et al.	Meta-Modelle	-	-	+	-	+
Universität Prag	XML-Schema	-	+	+	-	(+)
Christian Will	XML-Schema	+	+	+	-	(+)

+ : Kriterium betrachtet

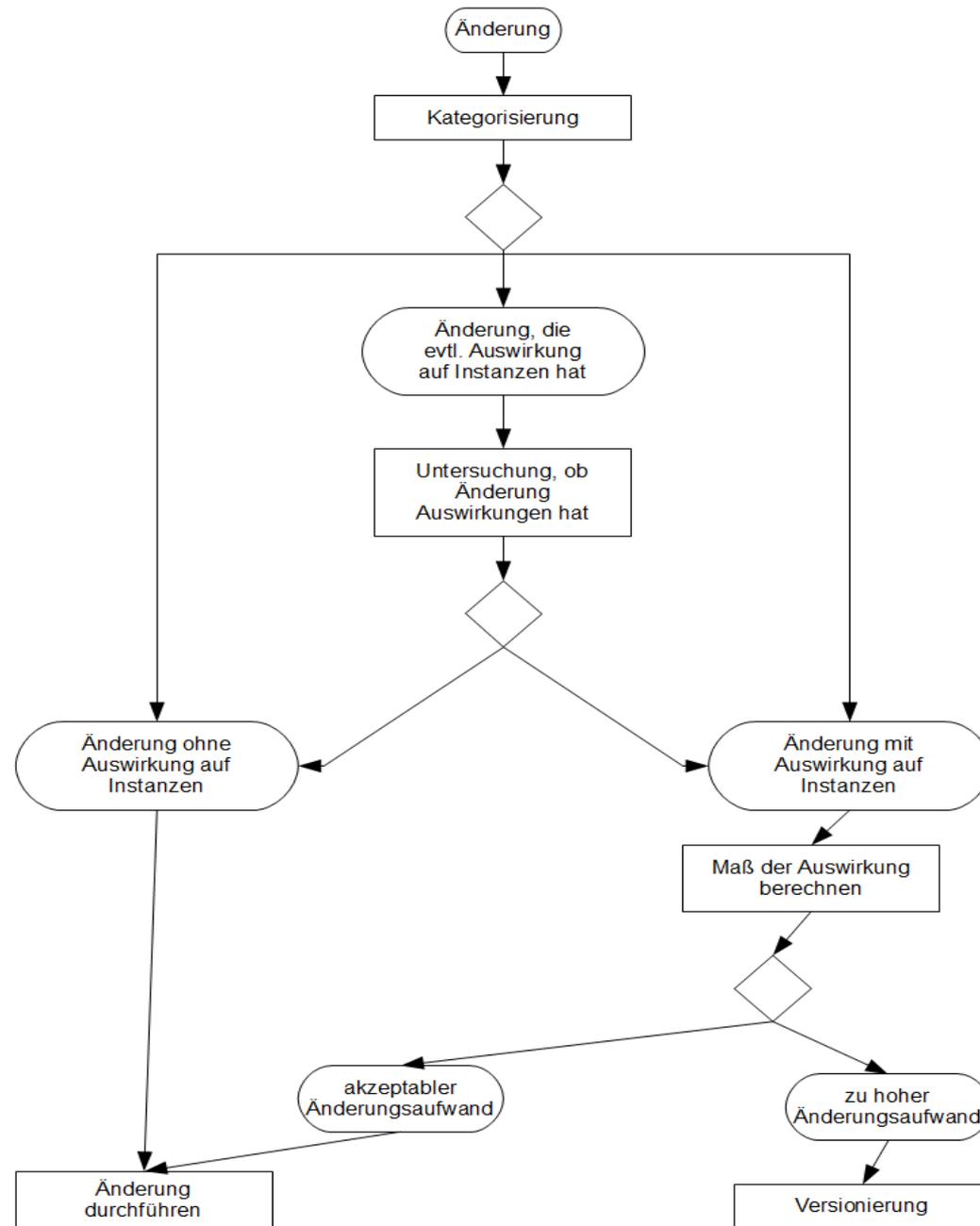
(+) : Kriterium teilweise betrachtet

- : Kriterium nicht betrachtet

Eigene Kategorisierung

- 1. Stufe: für Bewertung wichtig, ob die Änderung(en)
 - Keine Auswirkung auf Instanzen
 - Möglicherweise Auswirkung auf (einige) Instanzen
 - Auf jeden Fall Auswirkung auf Instanzen hat
- 2. Stufe: Art der Änderung
 - Änderung der Informationskapazität
 - Abschätzen, wie viele Informationen verloren gehen und neu eingegeben werden müssen
 - Update-Operationen teurer, da Hintereinanderausführung von Insert und Delete
- 3. Stufe: Was wird geändert?

Ablauf



Aufwand

- Anzahl betroffener Dokumente
- Anzahl der Knoten eines Dokumentes, die durch Insert-, Delete- und Update-Operationen verändert werden
 - absolut
 - prozentual
- Benutzereingriff
 - z.B. Einfügen neuer nicht-optionaler Elemente mit manueller Informationseingabe
- Informationsverlust
 - z.B. Löschen eines nicht-optionalen Elementes

Pseudocode Entscheidung

Eingabe: XML-Schema *xs*, Liste Evolutionsschritte *list*, maximale Kosten *max_cost*

Ausgabe: Evolution/Versionierung

```
current_cost := 0;  
for (schritt in list)  
    current_cost += calcCost(schritt, schema);  
    if(current_cost > max_cost)  
        return Versionierung;  
return Evolution;
```

Quellen

- [ACP09] Antonio Cicchetti, Davide Di Ruscio und Alfonso Pierantonio: Managing Dependent Changes in Coupled Evolution. In: ICMT 2009, Seiten 35 - 51, 2009.
- [GG05] Giovanna Guerrini, Marco Mesiti, Daniele Rossi: Impact of XML Schema Evolution on Valid Documents. In: WIDM 05, 2005.
- [GG06] Giovanna Guerrini, Marco Mesiti, Daniele Rossi: XML Schema Evolution, 2006.
- [Har07] Hartung, Michael: XML Schema Evolution. Technischer Bericht, Universität Leipzig - Abteilung Datenbanken, 2007.
- [HS01] Hong Su, Diane Kramer, Li Chen Kajal Claypool Elke A. Rundensteiner: XEM: Managing the Evolution of XML Documents. In: Proceedings of the 11th International Workshop on Research Issues in Data Engineering, 2001.
- [Mal10] Malý, Jakub: XML Schema Evolution. Diplomarbeit, Univerzita Karlova v Praze, 2010.
- [MMM07] Mirella M. Moro, Susan Malaika, Lipyeow Lim: Preserving XML Queries during Schema Evolution. In: WWW 2007, 2007.
- [MN09] Martin NeĚaský, Irena Mlýnková: Five-Level Multi-Application Schema Evolution. In: Dateso 2009, Seiten 90 - 104, 2009.
- [MT04] Marvin Tan, Angela Goh: Keeping Pace with Evolving XML-Based Specifications. In: EDBT 2004 Workshops, 2004.
- [PK05] Peter Kunzmann, Franz-Peter Burkard, Franz Wiedmann: dtv-Atlas Philosophie. Deutscher Taschenbuch Verlag, 2005.
- [Wil06] Will, Christian: Entwicklung und Implementierung einer Sprache zur Evolution von XML-Schemata. Diplomarbeit, Universität Rostock, 2006.
- [Zei01] Zeitz, Andre: Evolution von XML-Dokumenten. Studienarbeit, Universität Rostock, 2001.



Fragen?