
On classification of XML document transformations

Jana Dvořáková

FMFI UK, Bratislava

jana.dvorakova@dcs.fmph.uniba.sk

DATESO 2005

Outline

- Introduction
- Formalization of XML concepts
- Classification hierarchy
- Type transformations
- Formal models
- Conclusion

Introduction

- XML - popular standard for production of structured documents
 - document standards, data exchange between applications
- Various transformations needed
- Many transformation systems exist, based on different techniques
 - SynDoc, Scrimshaw, TRES, XSLT, Grif, Thot, VXT, CoST,...
- Aims
 1. to define classification of XML document transformations
 2. to examine relationships among defined groups of transformations

Formalization of XML concepts (1)

- XML document

- tree, where internal nodes are elements and leaves contain textual content

- Document Type

- context-free grammar
- valid XML documents - derivation trees

- Restrictions

- XML documents without attributes
- element names from finite and known set

Formalization of XML concepts (2)

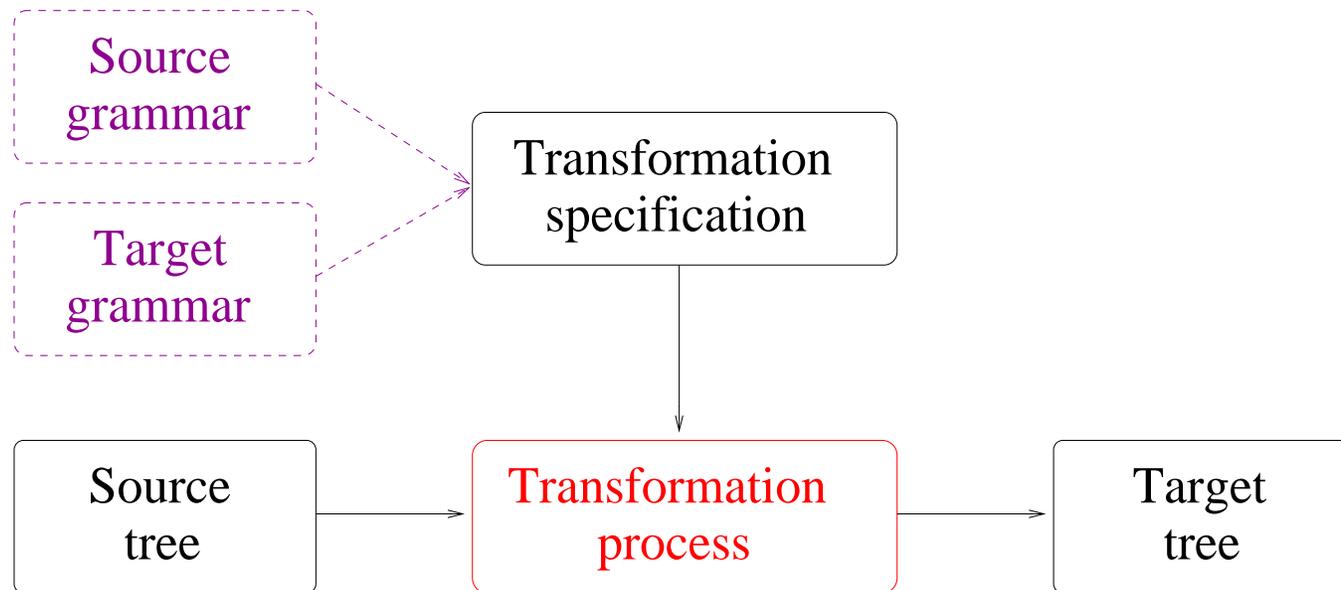
- Transformation

- set of pairs of trees (d_1, d_2)

d_1 - tree representing source XML document

d_2 - tree representing target XML document

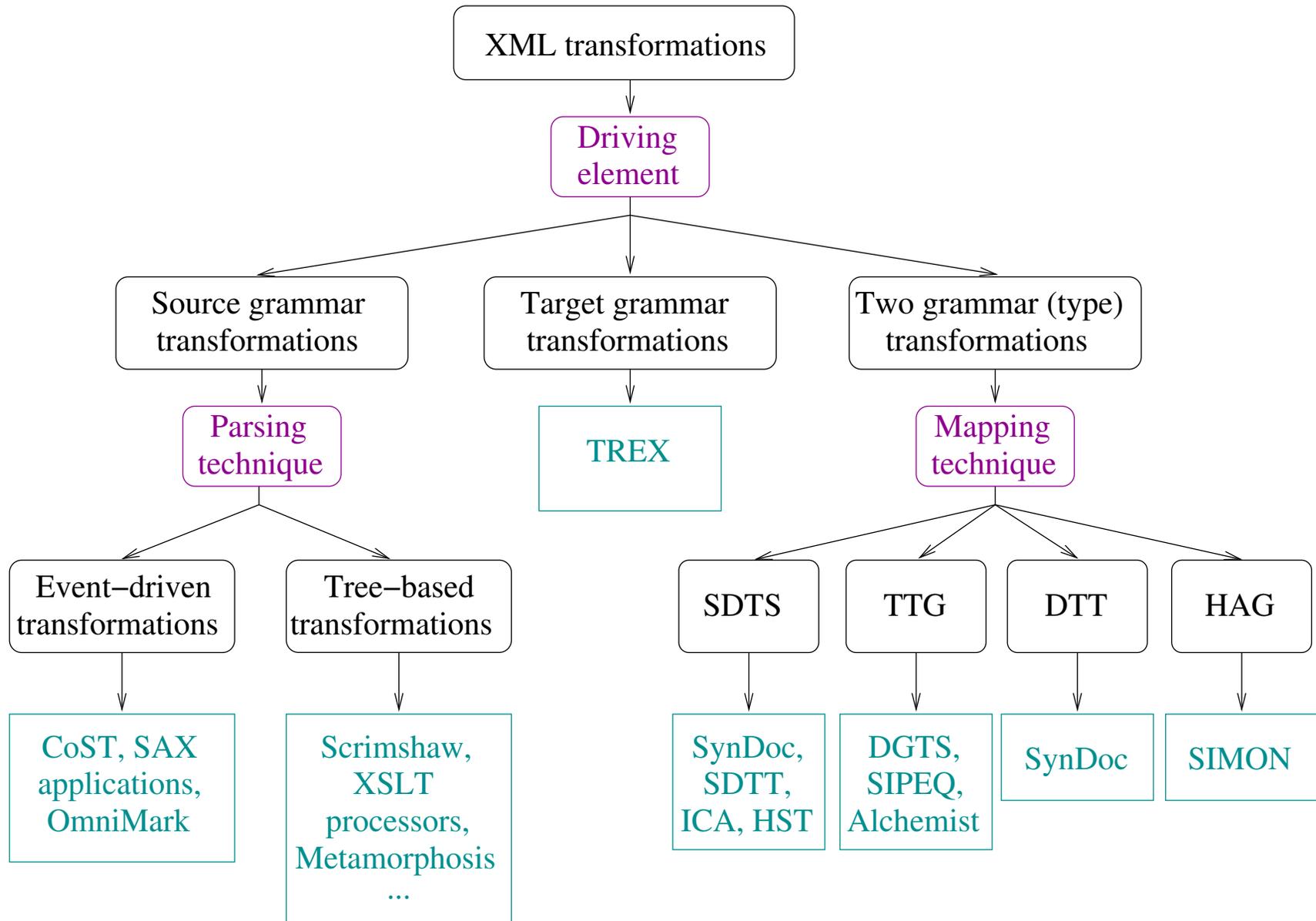
- Transformation model:



Classification hierarchy (1)

- Different possible criteria
 - driving element, scope, purpose, level of user interaction, transformational power, source and target correctness...
- Basic groups:
 - Source grammar transformations
 - Target grammar transformations
 - Two grammar (type) transformations
- In each group different transformation techniques used

Classification hierarchy (2)



Type transformations

- Transformation specification is created by grammar mapping
 - Both source and target correctness
 - Formal models used:
 - Syntax directed translation schema (SDTS)
 - Tree transformation grammar (TTG)
 - Descending tree transducer (DTT)
 - Higher order attribute grammar (HAG)
1. Common framework development
 2. Mutual comparison

Formal models - SDTS (1)

- Translates a sentence of a source grammar (w_s) into a sentence of a target grammar (w_t) according to translation groupings
- Simulates derivation of both sentences from the start symbol simultaneously
- Sentential form = current frontiers of the source tree and the target tree

Translation grouping:

$$A \rightarrow u_1 \dots u_n, v_1 \dots v_m$$

Translation:

$$(S, S) \Rightarrow \dots \Rightarrow (xAy, x'Ay') \Rightarrow (xu_1 \dots u_ny, x'v_1 \dots v_my') \Rightarrow \dots \Rightarrow (w_s, w_t)$$

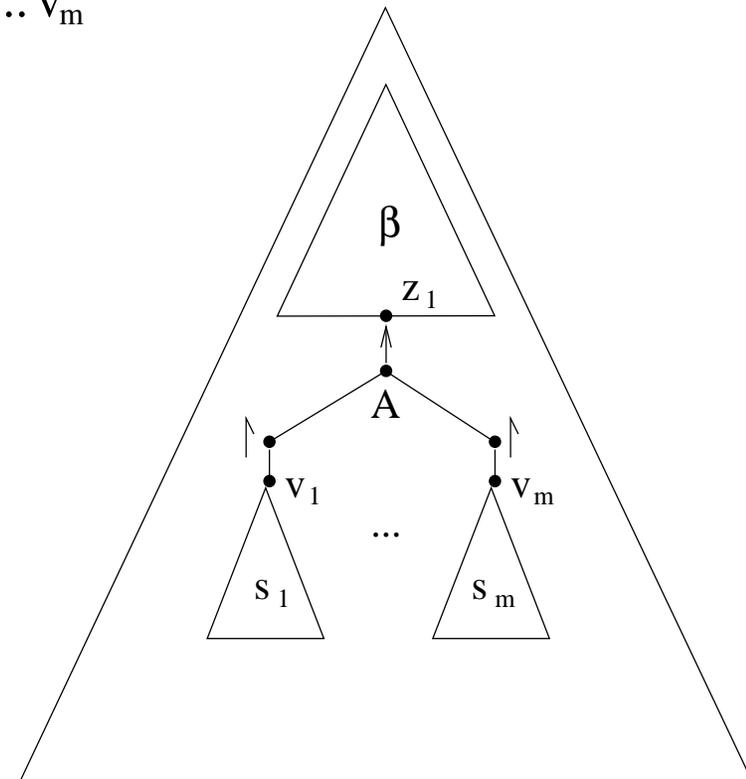
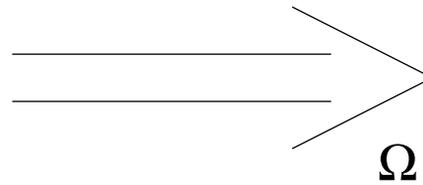
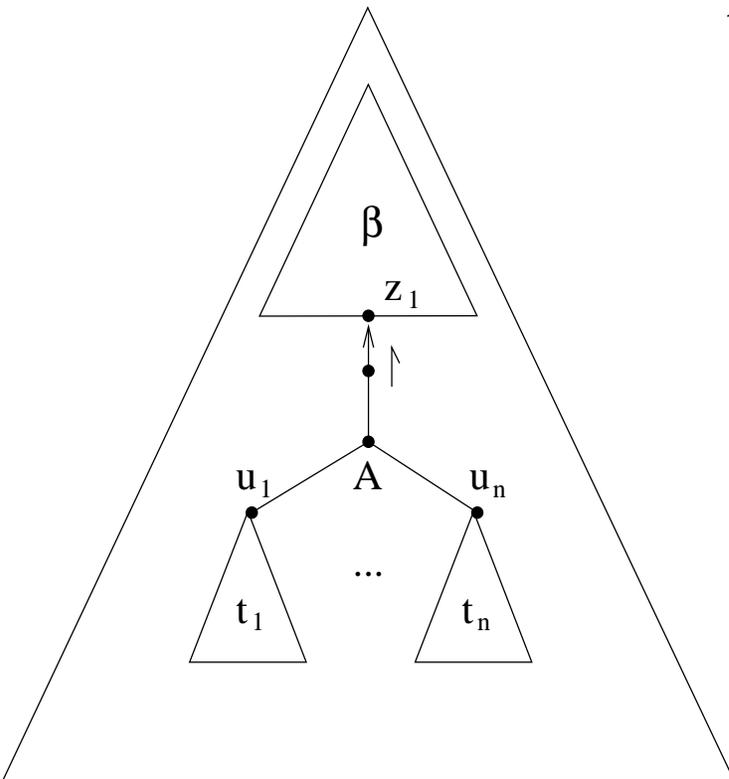
Formal models - SDTS (2)

- Standard definition is not suitable in the case of XML
- Extension proposed: trees as sequential forms
- First sequential form - source derivation tree, last sequential form - target derivation tree

Formal models - SDTS (3)

Translation step:

$$A \longrightarrow u_1 \dots u_n, v_1 \dots v_m$$



Formal models - comparison

- Results obtained by comparing syntax directed translation schema (SDTS), descending tree transducer (DTT) and their modifications:

	SDTS	ESDTS	d-DTT	DTT
SDTS		\subsetneq	$\not\subseteq$	\subsetneq
ESDTS	\supsetneq		N	N
d-DTT	$\not\subseteq$	N		\subseteq
DTT	\supsetneq	N	\supseteq	

Conclusion

- Summary of results:
 - formal definitions of XML concepts
 - classification for XML document transformations introduced
 - formal models for two grammar transformations defined in common framework
 - several results of comparing these formal models
- Future work:
 - add other comparisons of formal models
 - include attribute transformations

Thank you for your attention!