Grupo de Especificação e Processamento de Linguagens

Generating SGML specific editors
from DTDs to Attribute Grammars

José Carlos Ramalho
Alda Reis Lopes
Pedro Rangel Henriques

epl@di.uminho.pt

Grupo de Especificação e Processamento de Linguagens

Contents

- Behind the scene
- The main goal
- Attribute grammars: why and what
- DTD ⇒ AG conversion
- Future work

Grupo de Especificação e Processamento de Linguagens

Behind the scene

- Last year conference (Washington)
  - Semantic Validation: the possibility to add constraints and context conditions
- Markup Languages Journal
  - Processing constraints
    - type inference
    - value normalization

Grupo de Especificação e Processamento de Linguagens

Type inference

Problem: how to process ... ?

Constraint: **latitude > 39 and latitude < 42.5**

Document: **...<latitude>41.32</latitude>...**

Answer:

...<latitude type="float">41.32</latitude>...

Grupo de Especificação e Processamento de Linguagens

Value normalization

Problem: How can I identify ...?

... King <name>Affonso</name> proclaimed several ...

... And his soldiers battled against <name>Afonso</name>.

...and that church was built in the <date>XVIII century</date>.

...it all happened on <date> the fifth October</date>...

Answer

...King <name value="Afonso">Affonso</name>...

...it all happened on <date value="xxxx.10.05">the fifth...

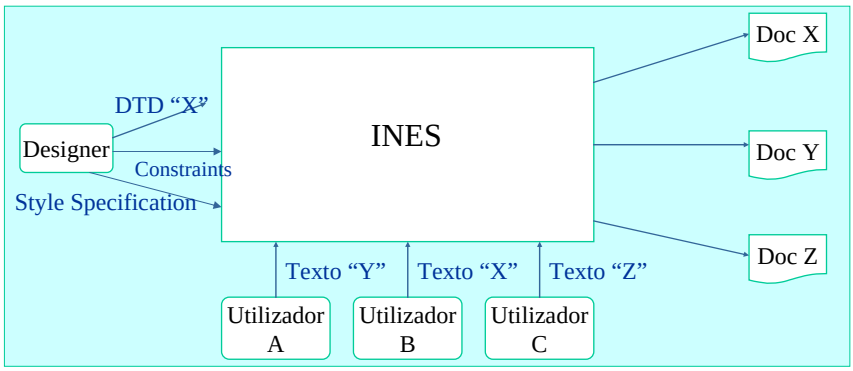
Grupo de Especificação e Processamento de Linguagens

Programs ⇔ SGML Documents

- Have a support language formally defined
- Processing - compiler
  - lexical analysis
  - syntactic analysis
  - semantic analysis
    - complex: type checking; type inference, ...
    - Can be formally specified: Attribute Grammars
- Have a support markup language defined in SGML
- Processing - parser
  - lexical analysis
  - syntactic analysis
  - semantic analysis
    - very simple: ID - IDREF coupling

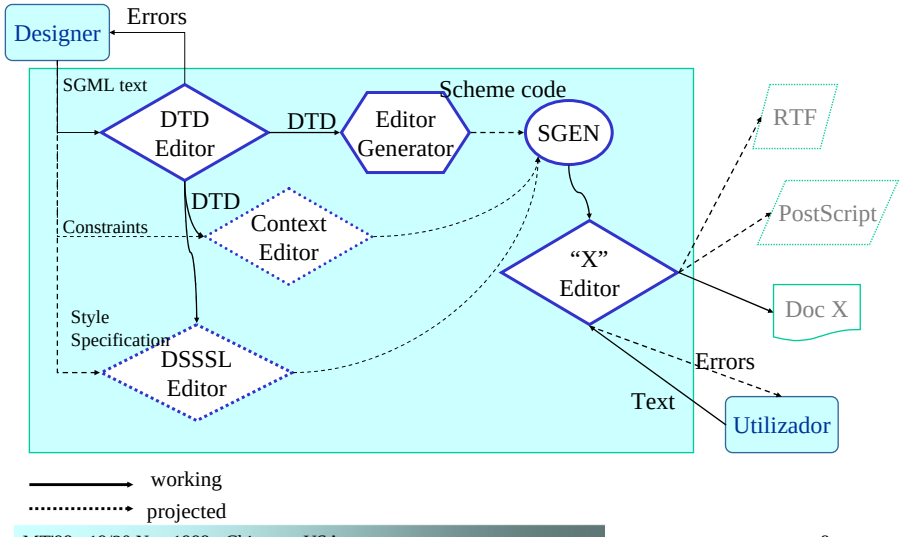
Grupo de Especificação e Processamento de Linguagens

INES - Document Programming Environment



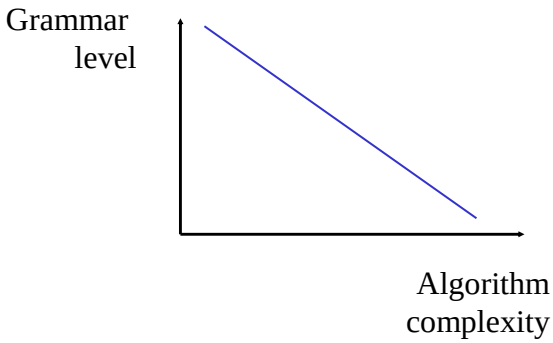
Grupo de Especificação e Processamento de Linguagens

INES: inside



Why Attribute Grammars?

The Chomsky Hierarchy



Grupo de Especificação e Processamento de Linguagens

Chomsky Hierarchy

Grammar level	Algorithm
• 0 - unrestricted	• Turing Machine
• 1 - Context-Sensitive	• Linear-Bounded Automaton
• 2 - Context-Free	• Push-Down Automaton
• 3 - Regular	• Finite-State Automaton

A grammar is classified by the highest Chomsky level it fits

Grupo de Especificação e Processamento de Linguagens

Attribute Grammars

To have a level 2 grammar
with level 1 expressive power!

Interesting languages have context!

- Is this element already defined?
- Are open brackets paired with closing ones?
- Does this variable have the correct type?

Grupo de Especificação e Processamento de Linguagens

Attribute Grammars

- Semantics
  - Context conditions and constraints as attribute equations
- New concepts in compiling
  - incremental parsing: being able to produce an instance of the AST at any time
  - incremental evaluation: recompiling only what is really necessary

AG: formal definition

AG = <G, A, R, C>

G is a context free grammar (level 2): G = <T, N, S, P>

- T - set of terminal symbols (alphabet)
- N - set of nonterminal symbols
- S - start symbol or axiom (S belongs to N)
- P - set of derivation rules

A is the set of all attributes: intrinsic, inherited and synthesized

R is the set of attribute evaluation rules

C is the set of all contextual conditions

AG: example

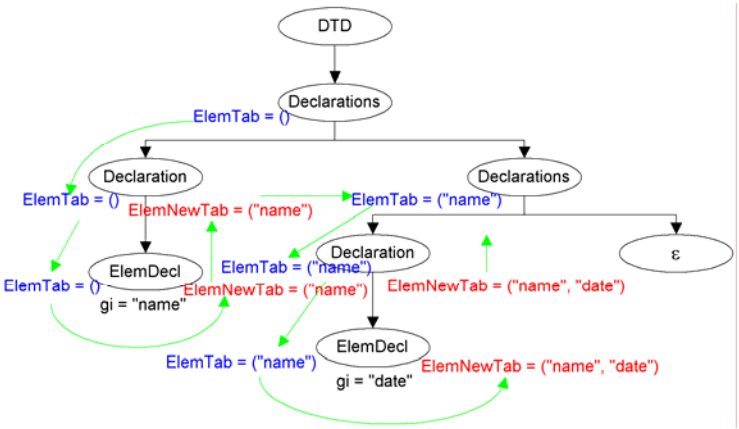
```
DTD → Decls
    Decls.ElemTab = ( )

Decls → Dec Decls
    Dec.ElemTab = Decls.ElemTab
    Decls$2.ElemTab = Dec.ElemNewTab
| ε

Dec → ElemDec
    ElemDec.ElemTab = Dec.ElemTab
    Dec.ElemNewTab = ElemDec.ElemNewTab
| AttDec ...

ElemDec  → gi min min Content
  if( not exist( gi, ElemDec.ElemTab ))
    ElemDec.ElemNewTab = insert( ElemDec.ElemTab, gi )
 else error("Element already defined!")
```

AG: example (cont.)



Grupo de Especificação e Processamento de Linguagens

AG implementation tool: SGen

- It allows attribute definitions
- It allows attribute equations definitions
- It provides an evaluator for those equations

- Sgen
  - based in a formal language: the interface, attribute equations, ...
  - multiple views of the internal representation

Grupo de Especificação e Processamento de Linguagens

DTD2AG conversion

- Elements and Attributes
- Content models
  - connectors
  - occurrence indicators
- SGML “specials”

The AG internal representation is very similar to a grove. Disjoint things in SGML will have to merge.

Grupo de Especificação e Processamento de Linguagens

Elements and Attributes

SGML:

```
<!ELEMENT Letter - - (Date,Name+,Message,End)>
<!ATTLIST Letter Type (work|family|friend) work
                 Idiom  (en|pt|fr|sp) en>
```

S1: Attributes as part of the content model
Lett

S2: Attributes apart from content model
Let
Let
Let

S3: Attributes as the AG synthesized attributes
Letter → Date NameList Message End
  Letter.type = ...
  Letter.idiom = ...

Grupo de Especificação e Processamento de Linguagens

Connectors

C1: ,

SGML:

<!ELEMENT Letter - - (Date,Name+,Message,End)>

AG:

Letter → Date NameList Message End

C2: |

SGML:

<!ELEMENT Letter - - (Date|Name+|Message|End)>

AG:

Letter → Date | NameList | ...

Grupo de Especificação e Processamento de Linguagens

Connectors

C3: &

SGML:

<!ELEMENT Letter - - (Date&Name+&Message&End)+>

AG1: all possible combinations

Letter → Date NameList ...

| NameList Date ...

...

AG2: transform SGML and add constraint

<!ELEMENT Letter - - (Date|Name+|Message|End)>

CC: ∀ x in Content, $\exists_1$x

Letter → X

Letter → X Letter

X → Date

if(exists(Date.gi, X.ElemTab)) then error "..."

| NameList

| Message | End

Grupo de Especificação e Processamento de Linguagens

Occurrence indicators

OI1: ?

SGML:

<!ELEMENT Letter - - (Date?|...)>

AG:

Letter → Date ...

Date → ε | ...

OI2: * or +

SGML:

<!ELEMENT Letter - - (Date|Message|...)*>

AG:

Letter → ε

| LetterContent Letter

Grupo de Especificação e Processamento de Linguagens

AG: an open path to semantics

Grupo de Especificação e Processamento de Linguagens

Questions?

Grupo de Especificação e Processamento de Linguagens

New SGML auth. and proc. model