Innsbruck University

Mini Seminararbeit

Data Langauges

Author:
Huda Alkazzaz

Supervisor:
Dr. Georg Moser

June 3, 2015

Contents

1 Introduction **1**

2 Data Language **1**
2.1 XML - eXtensible Markup Language 2
2.2 Page Description Languages(PDL) 3

3 Summary **3**

1 Introduction

Nowadays play the data language very important role in data format. In contrast to natural languages, most of artificial languages are programming languages. Grammars are the dominant means for language specification. Where as grammars are designed for virtually all programming languages, it is a pretty rare practice to use grammars for specifying data formats as opposed to computer programs. Normally, data formats are specified by giving interpretations to fixed-size chanks of the data. To understand the Grammars that specifying syntax,which is a branch of mathematics and computer science that formalizes the properties of "languages" over strings and their syntax.

2 Data Language

The idea behind language theory directly relevant to language design. Some terminology:

- A `language` is a set of strings over some alphabet. If Σ is an alphabet then Σ is also a language over Σ. So is Σ^*. Σ^* is important: it is the set of all strings over Σ. So for any language L over Σ, L $\subseteq \Sigma^*$. Also important is that Σ^* is countable.
- A `grammar` is a specification which (given an appropriate semantics) states which strings over some alphabet are in a language, and which are not. A grammar is formally defined as the tuple G = (V, Σ, P, S)consisting of:
  - V A finite amount, which is called vocabulary.
  - $\Sigma \subset$ V, A subset of V, it is called the alphabet and the elements are called terminal symbols.
  - P $\subset (V^* \setminus \Sigma^*) \times V^*$,A finite set of production rules, and
  - S $\in$ V $\setminus \Sigma$, The start symbol.
- A `parser` transforms a string in some language into a syntax tree that reflects the underlying of strings in that language. parser or syntactic analysis is the process of analysing a string of symbols, either in natural language or in computer languages, conforming to the rules of a formal grammar.

Usually there are formal tools for doing this as well:

- Backus-Naur `Form grammars` and regular expressions are usually used to specify syntax.
- Semantic `specifications` (which are considerably more complex) employs various formalisms — one of the more popular in academic language design is operational semantics. (Non-academic language designers tend to use ad hoc English specifications or reference implementations; these tend to be ambiguous or buggy respectively.
  The using of this grammar, which makes a file format of a data language. There are basically two groups of them. The first one is XML-related languages.In the view of XML advocates, XML-related languages eventually replace all other data languages and all other data formats for that matter. The second group combines all other data languages.It includes exclusively languages for printing and for visualization of documents.

2.1 XML - eXtensible Markup Language

The XML is a language that defines a set of rules for encoding documents in a format which is both human-readable and machine-readable. The XML is extensible and it serves as the basis for many descriptive languages and it is a subset of the Standard GeneralizedMarkup Language (SGML), a complex standardfor describing document structure and content.

It is a textual data format with strong support via Unicode for different human languages. Although the design of XML focuses on documents, it is widely used for the representation of arbitrary data structures such as those used in web services. The classic application of language found in parser generators or in general compiler construction .A parser generator converts the description of a language in a parser for that language.Implementations of parsers are available for all programming languages. different kind of systems exist to aid in the definition of XML-based languages, while many application programming interfaces (APIs) have been developed to aid the processing of XML data. XML is a tag based format as HTML, but it describes the content rather than the presentation of that content. The parser may translate the document into a tree in memory, accessible through to the Document Object Model (DOM). But you can also associate functions to tags Before to write an XML document, the operator should write a Document Type Declaration. A DTD declares a grammar of tags, and an XML document is an instance of that grammar as an object is an instance of a class, or as a program for a language. The DTD may be included into the XML document, or linked by an URL. Without the DTD, the XML document may be used but not checked for validity. A document is validated with:

- DCD (Document Content Description for XML). DCD is a language that provides a structural schema which replaces the functions of the DTD to describe constraints on tags and content of XML documents.
- A schema, as a DTD, describes the grammar of tags for validating XML documents.

XML is simple because its rules for creating amarkup language to encapsulate data are straight-forward.

2.2 Page Description Languages(PDL)

Printing languages is a high-level programming language that are representations of exactly what needs to be on the screen or printed page. They are generally a collection of drawing commands that programs can generate, often with extra features to make drawing complex pictures or doing fancy things with text easier.Data sent to a printer must be in a language that the printer can understand. These languages are called Page Description Languages, or PDLs language.Adobe's PostScript and Printer Control Language(PCL)are the two most commonly used PDLs:

- PostScript`program` is a set of instructions that draw the final document. Different fonts and graphics can be used.Usually this program is generated by application software, so the process is invisible to the user. PostScript is a very useful output form.
- Printer Command `Language`or PCL is an extension of ASCII, adding escape sequences for formatting, font selection, and printing graphics, to provide a generic printer language for their entire range of printers.It is more current incarnations are quite flexible and capable.

3 Summary

According of many articles the data language are programming languages. The grammars of this language are the language specification. Whereas grammars are designed for virtually all programming languages. Actually there two category of Programs. Firstly, XML is the most of widespread tool for data exchange and Web presentation. The advantage of XML derives from the fact that the aspects of structuring, representing and visualizing a piece of information are handled independently with specific tools. Secondly a document analysis method, which extracts layout and text information from document files of various formats. This kind of method calls the page description language (PDL) and analyzes data generated from a printed document. The converting of the document to PDL data, this method can handle various document formats. Graphic elements such as text objects, image objects, and path objects in the PDL data are analyzed to extract text and layout information (character size, character position, and table position).

References

[1] J. Shallit.*A Second Course in Formal Languages and Automata Theory,*,Cambridge University Press,(2009).

[2] J. Roy and A. Ramanujan.*Englisch,XML: data's universal language,* journals and industry magazines,(2000).

[3] E. Harper.*Speaking In Tongues: Sorting Out Variable Data Printing Languages,*Report: Analyzing Publishing Technologies,(2007).

[4] G. Deleuze.*Postscript on the Societies of Control,*, MIT Press, Cambridge, MA, pp. 3-7,(1992).