Antlr Unicode Preprocessor

v. 1.0-SNAPSHOT

Project Documentation

Jürgen

2011-07-27

Table Of Content

1 Home

------ Introduction ------ Simone Tripodi ------ 2007-2010

1.1 Introduction

The SPARQL query language for RDF is designed to meet the use cases and requirements identified by the RDF Data Access Working Group in RDF Data Access Use Cases and Requirements; detailed explanations can be found on the official W3C's SPARQL specifications

Purpose of this project is provide a cross-compiler ANTRL v3 grammar which is an implementation of the SPARQL grammar specifications.

1.2 Sub-project of the sparkle-g project

Preprocessor to convert Unicode literals into character values extending ANTLRFileStream.

The ANTLRUnicodePreprocessorFileStream class will convert unicode escape sequences into character values. The input stream is preprocessed before a SPARQL request is passed to the Sparkle grammar.

The preprocessor for replacing Unicode escape sequences '\uxxxx' with character values is implemented as a Finite State Automaton. Incomplete Unicode escape sequences are written back unchanged into the file stream. Optimizations to avoid write operations lead to MODIFIED_DATA_STATE and data_buffer_modified. The rationale for deviations from pure doctrine where to keep the number of status small. In the usual case of no Unicode escape sequences in the data stream the maxime is to do (almost) nothing, just loop in the START_STATE.