

Commercial Implementation of Text Recognition Tools for VLC

John Rausch
Senior Staff Consultant
Lexis-Nexis
john.rausch@lexis-nexis.com

Researchers typically use traditional lexical scanner and parser tools for projects requiring the recognition of complex text elements. The role relegated to the lexical scanner is usually the simple tokenization while the relationships of the tokens to one another is done by the parser. Implementation of products or features using these tools for VLC can require months of processing or even years.

In this talk, I will describe how extending the capabilities of the lexical scanner while optimizing its performance can allow it to complete the recognition work traditionally done by parsers. This technique allows for flexible reprocessing of VLC that might otherwise not be done when improvements to algorithms are made . I will illustrate this with a case study of the recognition of embedded case law citations, including anaphoric references, and case names.