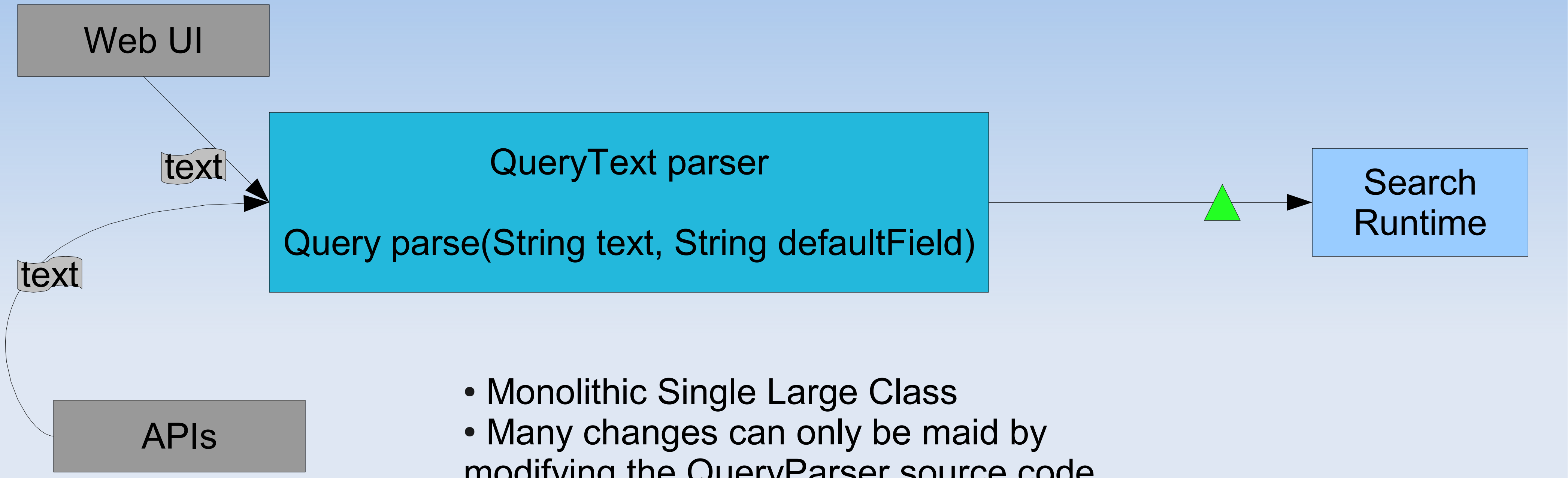
New QueryParser Framework for Lucene

Luis Alves
Adriano Campos
Michael Bush

June 2009

Old Lucene QueryParser



Web UI

text

text

QueryText parser

Query parse(String text, String defaultField)

Search Runtime

APIs

- Monolithic Single Large Class
- Many changes can only be maid by modifying the QueryParser source code
- Query logic, syntax and parsing mixed in to the same single class code.
- Difficult to maintain, support and add new features without breaking existing code.
- It's necessary to create multiple QueryParsers to support multiple syntaxes (there are multiple ones in contrib).

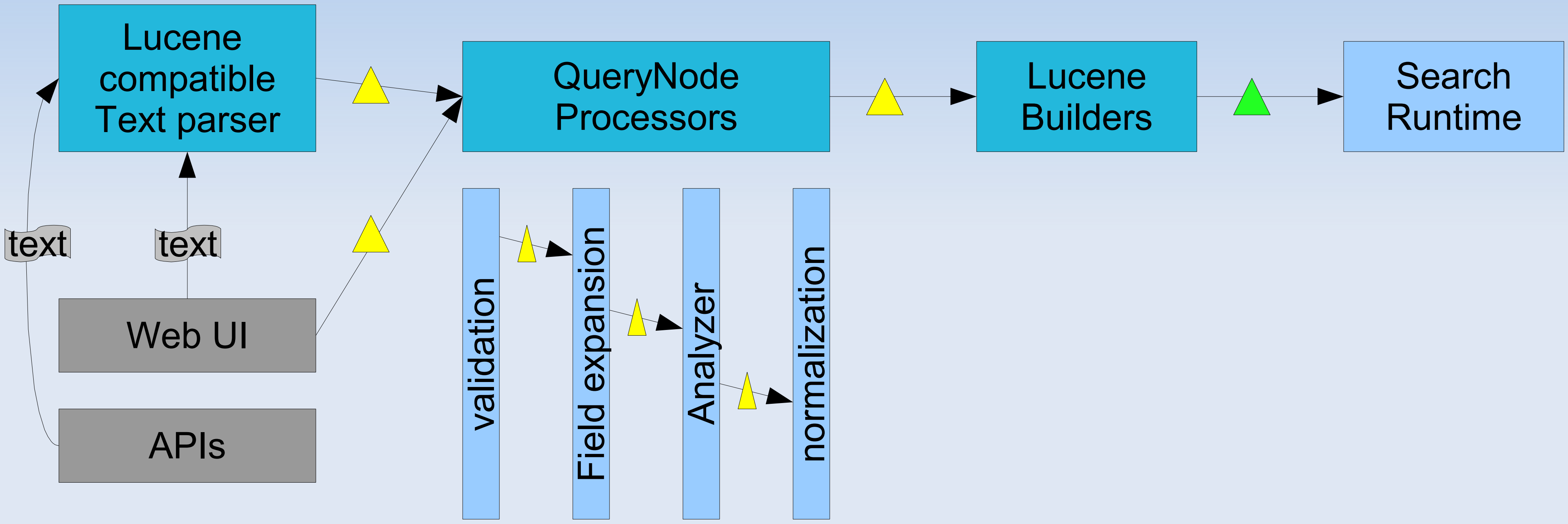
▲ - Lucene Tree

By Luis Alves - lafadev@gmail.com

New QueryParser features

- Parsing of text queries
- Syntax Tree optimization
- Syntax Tree expansion
- Syntax Tree validation and error reporting
- Tokenization and normalization of the query
- Makes it easy to create extensions
- Support for translation of error messages

New QueryParser framework



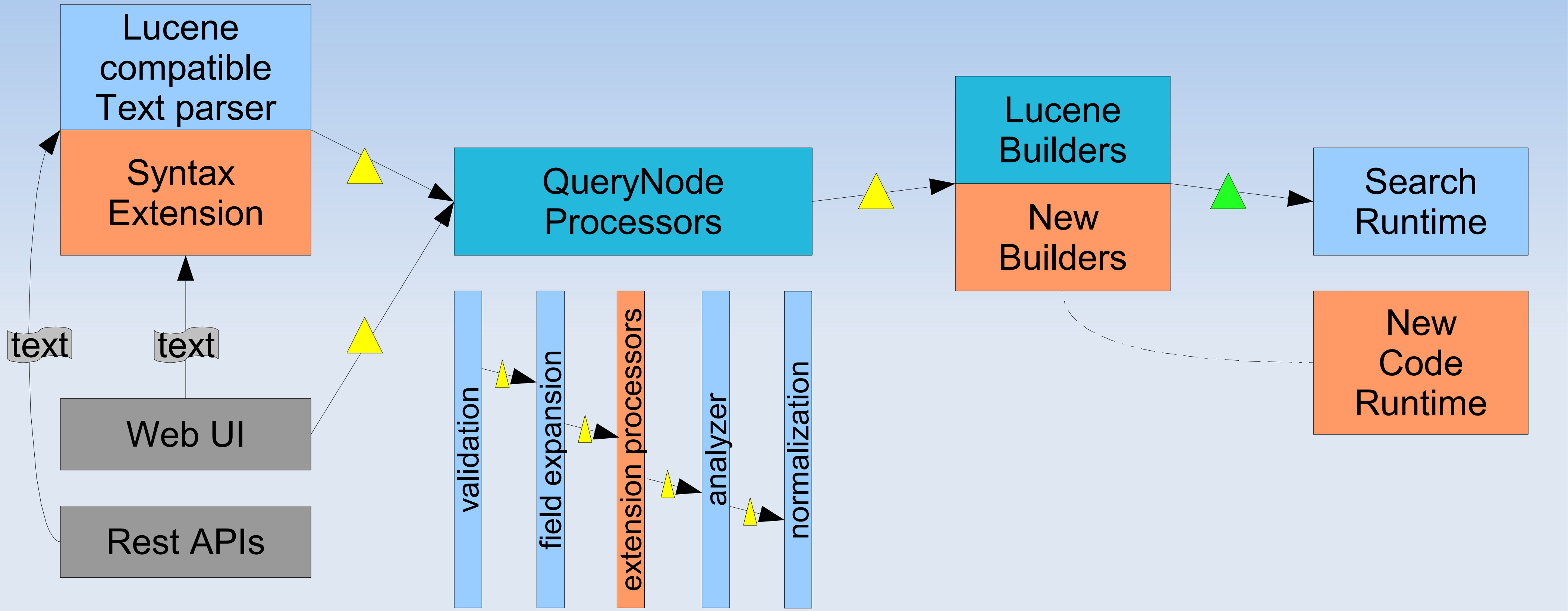
- 13 processors, few lines each
- builders for all Lucene Query Objects

🟨 - QueryNode Tree

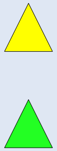
🟩 - Lucene Tree

By Luis Alves - lafadev@gmail.com

Extending new QueryParser



- Users can implementing a new parser that outputs a syntax tree
- Users can add new processors
- Users add new builders



By Luis Alves - lafadev@gmail.com

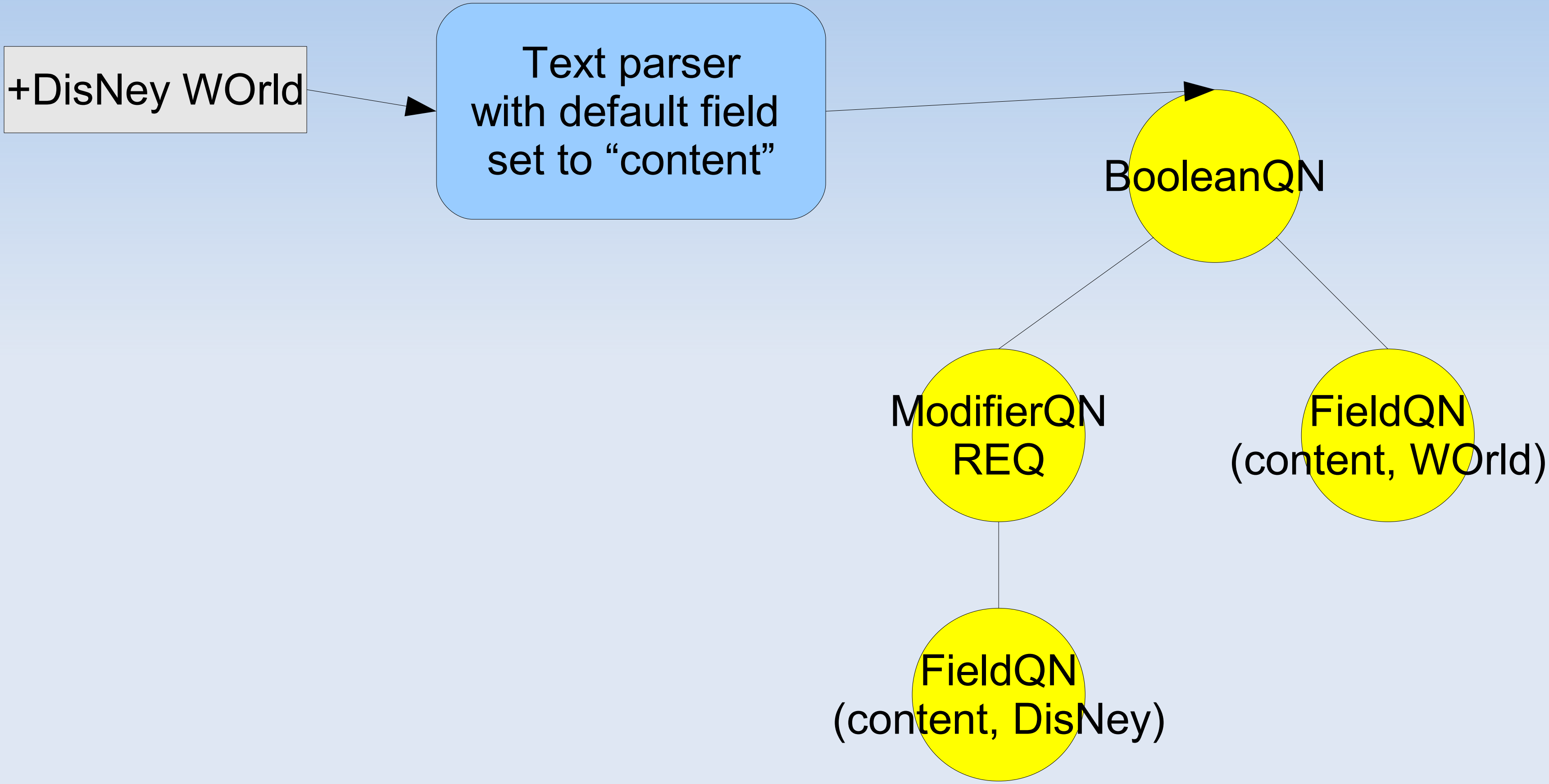
New QueryParser benefits

- Allows users to plug and play processors and builders, without having to modify lucene code.
- Allow lucene users to implement features much faster with less code.
- Allow users to change default behavior in a easy way without having to modify lucene code.

New QueryParser Example

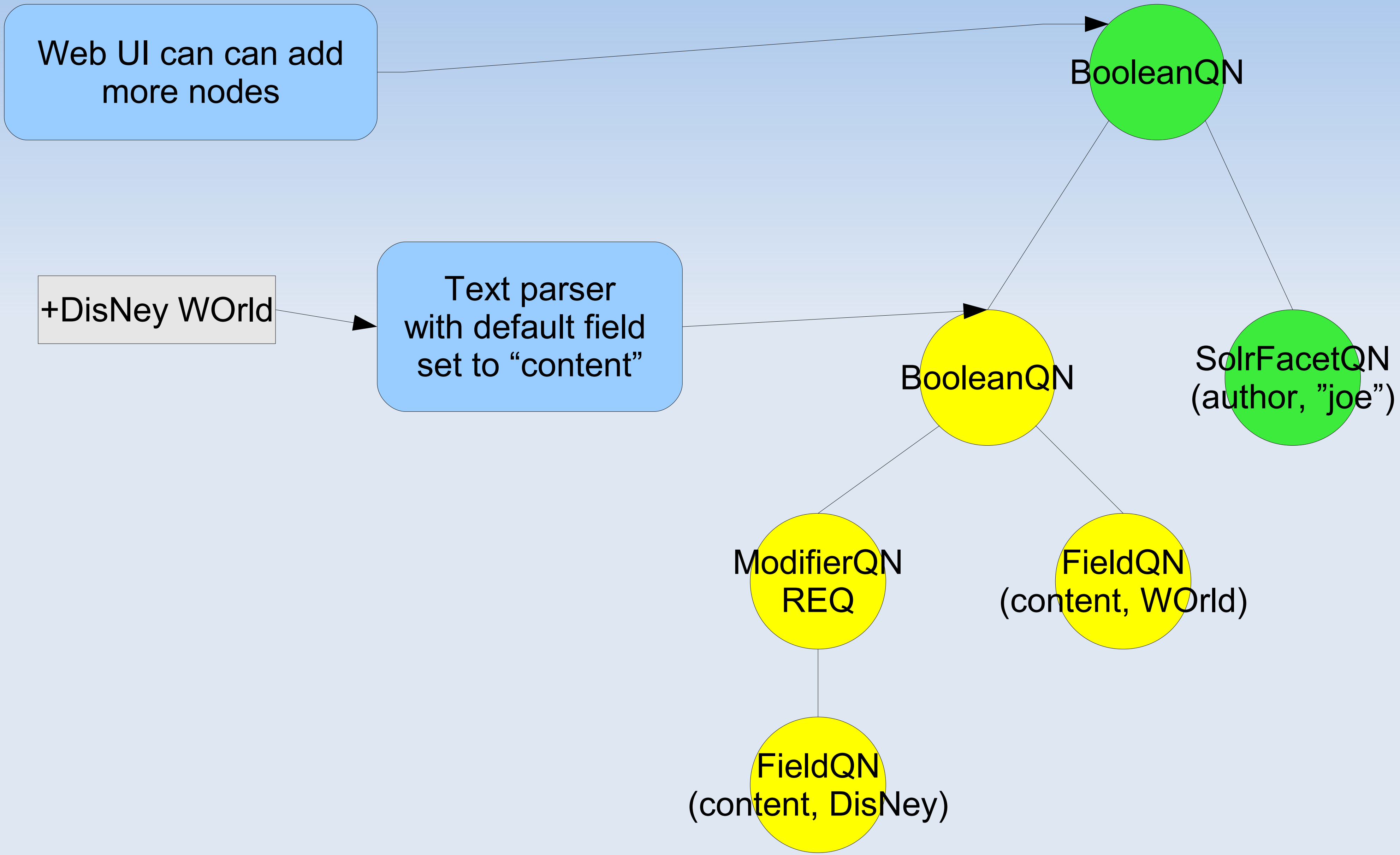
- Text parser creates a query-tree from a query-string
- The UI appends a simple field filter.
- A simple lowercasing Processor for field ”content”.

Text Parser



By Luis Alves - lafadev@gmail.com

UI + Text Parser



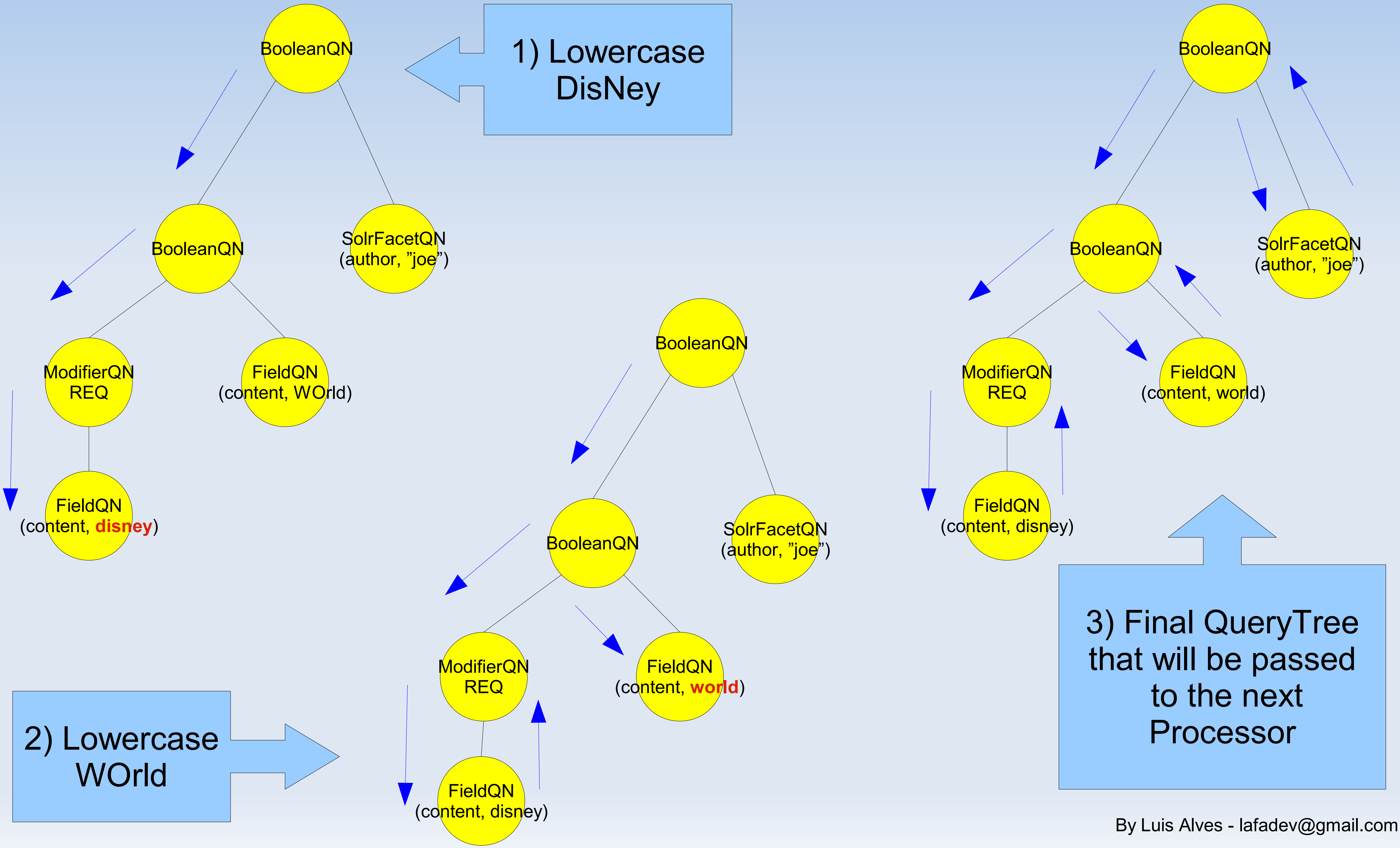
Web UI can can add more nodes

+DisNey WOrld

Text parser with default field set to “content”

By Luis Alves - lafadev@gmail.com

Lowercase Processor - example



1) Lowercase DisNey

2) Lowercase WOrld

3) Final QueryTree that will be passed to the next Processor

By Luis Alves - lafadev@gmail.com