

Indexing

When to create and delete indexes for STP and PA:

- when called iteratively on same input documents (as subexecutions): build index only once in parent execution;
- delete index at end of parent execution
 when called once as execution directly: build index in STP / PA directly; delete index at end
- current indexing:
- always:
- Bootstrappingon demand:SearchTermPosition
- PatternApplierif given as parameter:CommandLineInterpreter

Inconsistencies:

- CommandLineExecutor
 requires index-dir to be given for --search-candidates mode although SearchTermPosition can build index on demand in randomized temporary directory. User should be able to
 decide on the location of the index directory (by supplying index-dir) or to have a temporary index at a randomized directory (by not supplying index-dir).
 Proposed solution I: remove throwCLI statement when no index-dir is given in search-candidates mode
- (is specification I: remove throwcLI statement when no index-dir is given in search-candidates mode

 (is specification of an index dir even useful? The path to store temporary files can be adjusted using the config properties file…)

 Proposed solution II: remove index-dir parameter
- CommandLineExecutor & Bootstrapping
 if index-dir is given, CLI will create an index but Bootstrapping will build and use another temporary index.
 Proposed solution: add indexing on demand for bootstrapping

Temporary Files

created in:

- Indexer
 CommandLineExecuter (uses temporary datastore client for all sub-executions)
- Tagger
- InfolisBaseTestExecutionSchedulerTest

ApplyPatternAndResolveTest

- CommandLineExecuterTest
 ToytExtractorTest
- TextExtractorTestIndexerTestExampleChecker

Text Extraction

- on demand:
- RegexSearcher
- if given as parameter: CommandLineExecutor