contains the original document content plus any annotations added by the annotators. The document generator reads the CAS information and prepares the content for indexing by converting the document to a Lucene document.

Indexer

The *indexer* component is responsible for building a highly optimized index of document content that is suitable for high-speed text mining and analysis. The index is based on the open source *Apache Lucene indexer* with IBM extensions. IBM is an authorized committer of the Lucene open source project and frequently contributes selected features and functions back to the Lucene community.

When started, the indexer automatically indexes documents after they are processed by the document processors. You can manually perform a full rebuild of an index at any time. This option is useful if you add a text analytic to the UIMA pipeline and you want to include its results in your content analytics collection. If you enabled the document cache option for your collection, it is not necessary to recrawl your documents again. In this situation, the document content is obtained from the document cache, which is the temporary repository of crawled content.

Search run time

The search runtime component is a server-based component that is responsible for servicing user search and analytic requests. Client service requests are made by using the REST API. The REST APIs are a programming interface based on Java that operates remotely by using the HTTP/HTTPS protocol. The enterprise search application and content analytics miner are example REST APIs client applications that make service requests to a search runtime component.

A single content analytics collection is associated with at least one search run time or multiple search run times to support large-scale multiuser environments. The search runtime components are not dependent on the indexer component and are designed for continuous operation. To maintain this independence, they operate on copies of their associated content analytics collection. For more information about this topic, see 1.4.2, "Data flow" on page 24.

Content analytics miner

The content analytics miner is what you use to perform content analysis. Its user interface is browser-based and communicates with the content analytics miner that runs under either Jetty or WebSphere Application Server. Jetty is the default installation. The Content Analytics Miner web application issues REST client requests to the search run time associated with a given content analytics collection. The search runtime server component can either be installed locally or remotely from the analytics miner application.