

Collection Process

Publication Process

Sources

Pull
(Crawling)

Pull
(Harvesting)

Push
(SOAP/ReST)

Content Inbound

Syntactic

Semantic

Content Validation

Extraction

Enhancement

Filtering

Content Enrichment

Indexing

Actors

HTTP/Get
(URL)

HTTP/Post
(SOAP/ReST)

API
(Java, Perl,...)

Request Inbound

Syntactic

Semantic

Request Validation

Redirection
(Suggestions)

Enhancement
(add/remove clauses)

Request Enrichment

Filtering

Grouping

Sorting

Searching & Ordering

Search Engine

Stateless
(XSLT, SolrJS)

Statefull
(Webapp Framework)

Response Outbound

Redirection
(more like this)

Enhancement
(metadata, editorial)

Response Enrichment