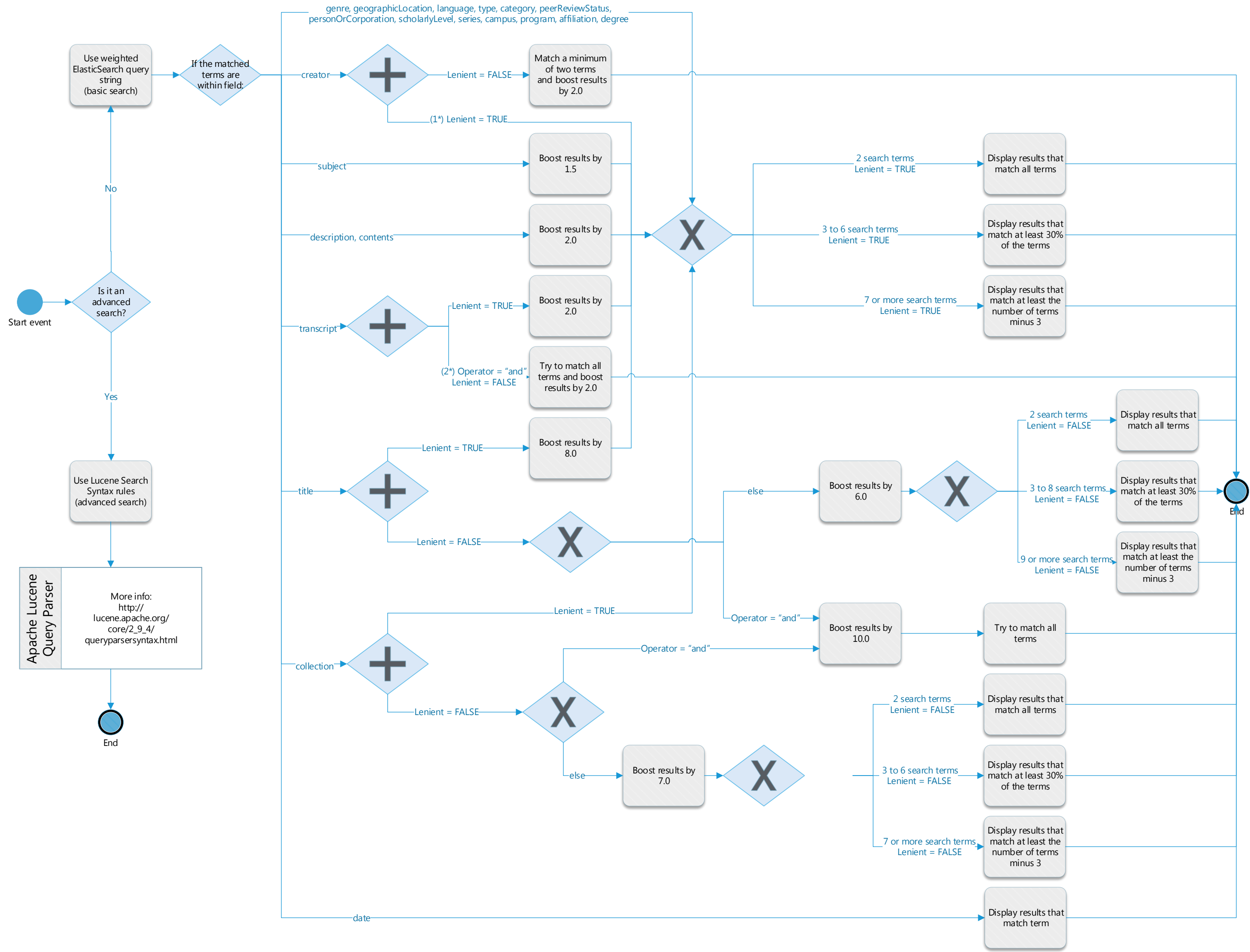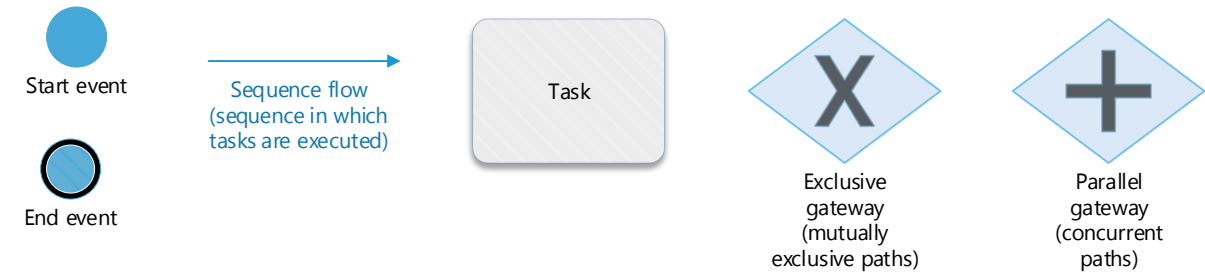# OC Search workflow - December 7th 2017

Business Process Model and Notation (BPMN) diagram depicting OC search workflow in Open Collections. Based on:
https://oc-index.library.ubc.ca/_queryString

Contact: Paul Joseph <paul.joseph@ubc.ca>, Carolina Roman Amigo <carolina.amigo@ubc.ca>

Start event

Is it an advanced search?

No

Use weighted ElasticSearch query string (basic search)

If the matched terms are within field:

creator

genre, geographicLocation, language, type, category, peerReviewStatus, personOrCorporation, scholarlyLevel, series, campus, program, affiliation, degree

Lenient = FALSE

Match a minimum of two terms and boost results by 2.0

(1*) Lenient = TRUE

subject — Boost results by 1.5

description, contents — Boost results by 2.0

transcript

Lenient = TRUE — Boost results by 2.0

(2*) Operator = "and" Lenient = FALSE — Try to match all terms and boost results by 2.0

2 search terms Lenient = TRUE — Display results that match all terms

3 to 6 search terms Lenient = TRUE — Display results that match at least 30% of the terms

7 or more search terms Lenient = TRUE — Display results that match at least the number of terms minus 3

Yes

Use Lucene Search Syntax rules (advanced search)

Apache Lucene Query Parser

More info:
http://lucene.apache.org/core/2_9_4/queryparsersyntax.html

End

title

Lenient = TRUE — Boost results by 8.0

Lenient = FALSE

else — Boost results by 6.0

2 search terms Lenient = FALSE — Display results that match all terms

3 to 8 search terms Lenient = FALSE — Display results that match at least 30% of the terms

9 or more search terms Lenient = FALSE — Display results that match at least the number of terms minus 3

End

collection

Lenient = TRUE

Operator = "and" — Boost results by 10.0 — Try to match all terms

Lenient = FALSE

else — Boost results by 7.0

2 search terms Lenient = FALSE — Display results that match all terms

3 to 6 search terms Lenient = FALSE — Display results that match at least 30% of the terms

7 or more search terms Lenient = FALSE — Display results that match at least the number of terms minus 3

date — Display results that match term

Legend

Start event

End event

Sequence flow (sequence in which tasks are executed)

Task

Exclusive gateway (mutually exclusive paths)

Parallel gateway (concurrent paths)

1* The lenient parameter can be set to true to ignore exceptions caused by data-type mismatches, such as trying to query a numeric field with a text query string.
(Source: https://www.elastic.co/guide/en/elasticsearch/reference/6.0/query-dsl-match-query.html)

2* The match query is of type boolean. It means that the text provided is analyzed and the analysis process constructs a boolean query from the provided text. The operator flag can be set to or or and to control the boolean clauses (defaults to or).
(Source: https://www.elastic.co/guide/en/elasticsearch/reference/6.0/query-dsl-match-query.html)