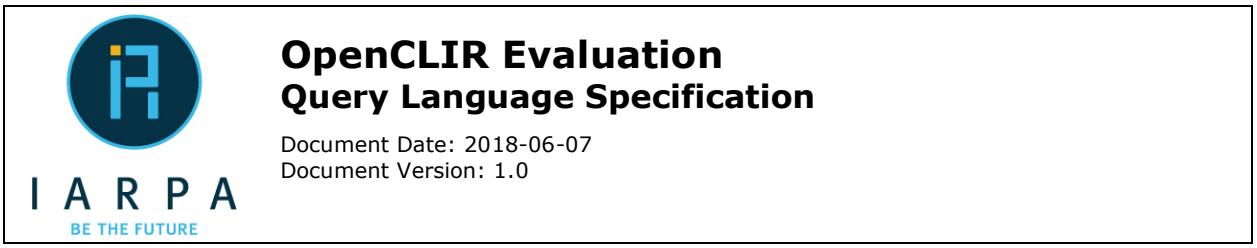

OpenCLIR Evaluation
Query Language Specification

Document Date: 2018-06-07
Document Version: 1.0

1 Overview

This document contains the formal specification for the Query Language for the OpenCLIR Evaluation[1] and the corresponding context-free grammar (CFG).

2 OpenCLIR Evaluation Query Language Specification

The table below describes syntax of various types of OpenCLIR queries. Section 3 below gives additional examples of allowed and disallowed queries.

Construct	Notes and Examples
x	x is a lexical query consisting of a single word. e.g., dictator
"x y"	x y is a single, multi-word term and is to be treated as a lexical query (as opposed to two separate terms subject to conjunction). e.g., "social media"
x+	x is conceptual and is subject to semantic expansion. e.g., beekeeping+
"x y"+	x y is a single, multi-word term subject to semantic expansion. e.g., "climate change"+
EXAMPLE_OF(x)	x is a single concept (consisting of one or more words) subject to limited semantic expansion. Subtypes and instances of x are relevant to this query; other topically related terms are not relevant. x is NOT relevant to EXAMPLE_OF(x). e.g., ‘cobra’ is relevant EXAMPLE_OF(snake) but ‘snake’ is not e.g., EXAMPLE_OF(subatomic particle)
"x EXAMPLE_OF(y)" "EXAMPLE_OF(y) x"	x is treated as lexical; y is subject to limited semantic expansion as described for EXAMPLE_OF(x) above. e.g., `yellow daisy’ and `yellow lily’ are relevant to "yellow EXAMPLE_OF(flower)" but `dandelion’ and `yellow flower’ are not e.g., "EXAMPLE_OF(sport) player"
"(x y) z" "w (x y)"	x y is a constituent phrase within a larger phrase. Only one level of parentheses and one constituent per query are allowed. e.g., "(social media) post" e.g., "paintings of (people on trains)"
	Constituent phrases cannot appear in the same term as an EXAMPLE_OF() construct, or within a semantic or morphological constraint (see table of examples in Section 3). e.g., the following is disallowed: EXAMPLE_OF((social media) post) e.g., the following is disallowed: "(paintings of people) on EXAMPLE_OF(vehicle)" e.g., the following is disallowed: abdicate [syn: (give up) power]
x, y	x and y are separate terms subject to conjunction. The order of the two terms is unimportant. Conjunction of two conceptual terms is disallowed, at least in the base period. Conjunction of one conceptual term and one lexical term is allowed (and constitutes a hybrid query). Conjunction of more than two terms is disallowed.

[1] https://www.nist.gov/itl/iad/mig/openclir-evaluation

	e.g., oil, tax e.g., vaccination+, autism e.g,. habitat, EXAMPLE_OF(rodent)
x [t: y]	y is a semantic constraint of type t on the query term x. Valid values for t are "hyp" (hypernym), "syn" (synonym), and "evf" (event frame). Semantic constraints can apply to individual words (including individual words inside quoted phrases), to phrases in quotes, or to EXAMPLE_OF() constructs. When applied to a term subject to semantic expansion, a constraint follows the plus sign. e.g., retreat [syn: withdraw] e.g., "arrive at the bank [evf: finance]" e.g., "golf club" [hyp: organization] e.g., strike+ [evf: labor] e.g., EXAMPLE_OF(virus) [evf: medicine] e.g., back [evf: anatomy], sore [syn: aching]
	At most one constraint per query term is allowed. e.g., the following is disallowed: nurse [hyp: profession] [evf: medicine] e.g., the following is disallowed: "guard [syn: protector] tower [hyp: structure]" e.g., the following is disallowed: "guard [syn: protector] tower" [hyp: structure]
	Semantic constraints cannot occur within parenthesized constituents or be applied to parenthesized constituents inside a larger quoted phrase. e.g., query strings such as "(a [t: x] b) c" and "(a b)[t: x] c" are disallowed
	Semantic constraints also cannot occur within an EXAMPLE_OF() construct or within a morphological constraint (see table of examples in Section 3). e.g., EXAMPLE_OF(a [t: x]) is disallowed e.g., <a [t: x]> is disallowed
<x>	x is a term that is subject to morphological constraint. Only lexical (non-conceptual) terms may be morphologically constrained. e.g., <contaminated>, water
	Morphological constraints cannot occur in the same query term as a parenthesized constituent phrase, an EXAMPLE_OF() construct, or a semantic constraint. e.g., the following is disallowed: "(paintings of <people>) on trains" e.g., the following is disallowed: "(paintings of people) on <trains>" e.g., the following is disallowed: EXAMPLE_OF(<mammals>) e.g., the following is disallowed: fish [hyp: <animals>]

3 Examples of Allowed and Disallowed Constructs

The table below shows examples of allowed and disallowed queries many of which are complex queries that combine multiple query types listed in Section 2[2].

Query	**Allowed/Disallowed**	**Why disallowed**
Lexical		
word	Allowed	
"word"	Disallowed	Extra quotes
"word word"	Allowed	
word word	Disallowed	Missing quotes or conjunction
"word word word"	Allowed	
Conceptual		

[2] See Section 5 for a note regarding spaces in examples.

word+	Allowed	
word+,word+	Disallowed	More than one conceptual term
EXAMPLE_OF		
"red EXAMPLE_OF(attire)"	Allowed	
"small green EXAMPLE_OF(vegetable)"	Allowed	
"EXAMPLE_OF(attire) word"	Allowed	
"EXAMPLE_OF(attire) word word"	Allowed	
Constituents		
"(word word) word"	Allowed	
"word (word word)"	Allowed	
"word (word word) word"	Allowed	
"word(word word) word"	Disallowed	Space missing
(word)	Disallowed	Constituent not within larger phrase
(word word)	Disallowed	Constituent not within larger phrase
"(social media) posts about paintings of (people on trains)"	Disallowed	Multiple constituents
"(word word)"	Disallowed	Constituent not within larger phrase
"posts about (paintings of (people on trains))"	Disallowed	Nested constituent
Conjunction		
word,word	Allowed	
"word,word"	Disallowed	Extra quotes
Semantic Constraints		
word[evf:test]	Allowed	
"word"[evf:test]	Disallowed	Extra quotes
"word[evf:test]"	Disallowed	Extra quotes
"word word"[evf:test]	Allowed	
"word word[evf:test]"	Allowed	
Morphological Constraints		
<contaminated>	Allowed	
"<contaminated>"	Disallowed	Extra quotes
"<contaminated> water"	Allowed	
<contaminated water>	Allowed	

"<contaminated water>"	Disallowed	Extra quotes
"<packets> of tea"	Allowed	
<packets of tea>	Allowed	
"<woke up> early"	Allowed	
"sheep <will jump>"	Allowed	
"<will pay off> debt"	Allowed	
Combined Types		
word+[syn:word]	Allowed	
"(word word) <word> word word"	Allowed	
"(word word) <word> word[evf:test]"	Allowed	
"(word word) <word>"	Allowed	
"(word word) word[evf:test] <word>"	Allowed	
"(word word) word[evf:test]"	Allowed	
"<contaminated> water"	Allowed	
"<contaminated> water"[evf:test]	Allowed	
"<word> (word word) word[syn:word]"	Allowed	
"<word> (word word)"	Allowed	
"<word> word"	Allowed	
"<word> word[evf:test] (word word)"	Allowed	
"<word> word[evf:word]"	Allowed	
"word <word> (word word) word"[evf:test]	Allowed	
"word <word> (word word) word[evf:test]"	Allowed	
"word <word>"	Allowed	
"word word[hyp:word] word (word word)"	Allowed	
"word[evf:test] (word word) <word>"	Allowed	
"word[evf:test] (word word)"	Allowed	
"word[evf:test] <word> (word word)"	Allowed	
"word[evf:test] <word>"	Allowed	
"word[evf:word] <word> word"	Allowed	

<contaminated water>[evf:test]	Allowed	
"red EXAMPLE_OF(attire)[evf:clothing]"	Allowed	
"(social media) posts[syn:publication]"	Allowed	
"<paintings> of (people on trains)"	Allowed	
virus[evf:medicine]	Allowed	
virus+[evf:medicine]	Allowed	
post[syn:publication], virus[evf:medicine]	Allowed	
"yellow EXAMPLE_OF(flower)[syn:blossom]"	Allowed	
"EXAMPLE_OF(attire) word"	Allowed	
"EXAMPLE_OF(attire)[evf:clothing] word"	Allowed	
"EXAMPLE_OF(vegetable) word word"	Allowed	
<word word+>	Disallowed	Morphological constraint of a conceptual term
<word word>+	Disallowed	Semantic expansion of morphological constraint
<word+>	Disallowed	Morphological constraint of a conceptual term
<word>+	Disallowed	Semantic expansion of morphological constraint
"[hyp:word] word"	Disallowed	Unclear what is being constrained
"(<word> word) word"	Disallowed	Constituent cannot be morphologically constrained
"(word+ word) word"	Disallowed	Conceptual term as a contituent
"<contaminated> water"+	Disallowed	Morphologically constrained conceptual term
"paintings of (people on trains)[syn:passengers]"	Disallowed	Constituent phrase can't be semantically constrained
"paintings of (people on trains[hyp:transportation])"	Disallowed	Constituent cannot be semantically constrained
"word[syn:word]"+	Disallowed	Extra quotes; constraint precedes plus sign
(word+ word)	Disallowed	Conceptual term as a contituent

(word+)	Disallowed	Conceptual term as a contituent
<fly[hyp:insect]>	Disallowed	Cannot semantically constrain a component of a morphologically costrained query
<word (word word)>	Disallowed	Cannot have a constituent phrase as part of a morphologically costrained query
<word[hyp:word] word>	Disallowed	Cannot semantically constrain a component of a morphologically costrained query
EXAMPLE_OF(<word> word)	Disallowed	Cannot morphologically constrain an E_OF concept
EXAMPLE_OF(word (word word))	Disallowed	Cannot have a constituent phrase as part of an E_OF concept
EXAMPLE_OF(word[hyp:test] word)	Disallowed	Cannot semantically constrain an E_OF concept
virus[evf:medicine]+[hyp:infectious agent]	Disallowed	Cannot have more than one semantic constraint
"<red> EXAMPLE_OF(attire)"	Disallowed	Cannot combine morphological constraints with EXAMPLE_OF
"red EXAMPLE_OF(attire)"[evf:clothing]	Disallowed	Cannot combine semantic constraint and EXAMPLE_OF
"(small green) EXAMPLE_OF(vegetable)"	Disallowed	Cannot combine constituent phrases with EXAMPLE_OF
"<will drink> <contaminated> water"	Disallowed	Multiple morphological constraints
virus[evf:medicine]+	Disallowed	Constraint precedes plus sign
"word[syn:word] word" [evf:word]	Disallowed	Cannot have more than one semantic constraint
"EXAMPLE_OF(attire) <red>"	Disallowed	Cannot combine morphological constraints with EXAMPLE_OF
"EXAMPLE_OF(attire) red"[evf:clothing]	Disallowed	Cannot combine semantic constraint and EXAMPLE_OF
"EXAMPLE_OF(vegetable) (word word)"	Disallowed	Cannot combine constituent phrases with EXAMPLE_OF

4 The MQL Parser Grammar

The following is the CFG for the parser. The comments are included to aid readability; they are not part of the formal specification.

```
####################################
# OpenCLIR Query Language (MQL) CFG
# Version: 1.0
# Date: 2018-06-07
####################################
# v1.2.1 modifies the extended EXAMPLE_OF form
#
####################################
# v1.2 adds an extended EXAMPLE_OF form
#
####################################
# v1.1 implements the following restrictions on queries.
# In the case of conjunctive queries, the restrictions apply to each conjunct
separately.
#
#  - At most one semantic constraint per component
#  - At most one morphological constraint per component
#  - At most one constituent per component
#
# In addition:
#
#  - Morphological constraints are not allowed in conceptual components
#  - Neither semantic constraints nor morphological constraints are allowed in
constituents
#  - Constituents cannot be semantically constrained
#
# And semantically constrained single-word conceptual query components
# must have the semantic constraint after the concept maker (plus sign).
#
# So the conceptual query form:
#     word[type:word]+
# is no longer allowed. The correct form is:
#     word+[type:word]
#
####################################
OPENCLIR_QUERY -> LEXICAL_COMPONENT | CONCEPTUAL_COMPONENT | CONJUNCTION
```

```
#-----------------------------------

# Define CONJUNCTION

# Only two conjuncts; cannot both be conceptual; order independent

#-----------------------------------

CONJUNCTION -> LEXICAL_COMPONENT "," LEXICAL_COMPONENT | LEXICAL_COMPONENT ","
CONCEPTUAL_COMPONENT | CONCEPTUAL_COMPONENT "," LEXICAL_COMPONENT

#-----------------------------------

# Define CONCEPTUAL_COMPONENT

#-----------------------------------

CONCEPTUAL_COMPONENT -> SIMPLE_CONCEPTUAL | EXAMPLE_OF | EXAMPLE_OF CONSTRAINT |
EXTENDED_EXAMPLE_OF

EXTENDED_EXAMPLE_OF  -> '"' WORD_OR_WORDS S EXAMPLE_OF '"' | '"' WORD_OR_WORDS S
EXAMPLE_OF CONSTRAINT '"' | '"' EXAMPLE_OF S WORD_OR_WORDS '"' | '"' EXAMPLE_OF
CONSTRAINT S WORD_OR_WORDS '"'

EXAMPLE_OF -> EX "(" WORD_OR_WORDS ")"

SIMPLE_CONCEPTUAL -> '"' ONE_CONSTITUENT_TERM '"' '+' CONSTRAINT | '"'
MORPHLESS_MAYBE_CONSTRAINED_TERM_CONTENT '"' '+' | '"' WORDS '"' '+' CONSTRAINT | WORD
'+' CONSTRAINT | WORD '+'

MORPHLESS_MAYBE_CONSTRAINED_TERM_CONTENT -> CONSTITUENT_AND_CONSTRAINT |
ONE_CONSTRAINT_TERM | ONE_CONSTITUENT_TERM | WORDS

#-----------------------------------

# Define LEXICAL_COMPONENT

#-----------------------------------

LEXICAL_COMPONENT -> '"' MAYBE_CONSTRAINED_TERM_CONTENT '"' | '"'
UNCONSTRAINED_TERM_CONTENT '"' CONSTRAINT | WOM | WOM CONSTRAINT

MAYBE_CONSTRAINED_TERM_CONTENT -> MORPH_AND_CONSTITUENT_AND_CONSTRAINT |
MORPH_AND_CONSTRAINT | CONSTITUENT_AND_CONSTRAINT | MORPH_AND_CONSTITUENT |
ONE_MORPH_TERM | ONE_CONSTITUENT_TERM | ONE_CONSTRAINT_TERM | WORDS

UNCONSTRAINED_TERM_CONTENT -> MORPH_AND_CONSTITUENT | ONE_MORPH_TERM |
ONE_CONSTITUENT_TERM | WORDS

WOM -> WORD | MORPH

#-----------------------------------

# Build up all combinations of term content involving at most one each of MORPH,
CONSTITUENT, and CONSTRAINT

#-----------------------------------

# Exactly three

MORPH_AND_CONSTITUENT_AND_CONSTRAINT -> MORPH_AND_CONSTITUENT ONE_CONSTRAINT_SUFFIX |
MORPH_AND_CONSTRAINT S ONE_CONSTITUENT | CONSTITUENT_AND_CONSTRAINT S ONE_MORPH

# Exactly two

MORPH_AND_CONSTITUENT -> ONE_MORPH S ONE_CONSTITUENT | ONE_CONSTITUENT S ONE_MORPH

MORPH_AND_CONSTRAINT -> ONE_MORPH ONE_CONSTRAINT_SUFFIX | WORD_OR_WORDS CONSTRAINT S
ONE_MORPH

CONSTITUENT_AND_CONSTRAINT -> ONE_CONSTITUENT_PREFIX ONE_CONSTRAINT_SUFFIX |
WORD_OR_WORDS CONSTRAINT S ONE_CONSTITUENT
```

```
# Exactly one

ONE_MORPH -> ONE_MORPH_TERM | MORPH

ONE_MORPH_TERM -> MORPH S WORD_OR_WORDS | WORD_OR_WORDS S MORPH | WORD_OR_WORDS S
MORPH S WORD_OR_WORDS

ONE_CONSTRAINT_SUFFIX -> CONSTRAINT S WORD_OR_WORDS | CONSTRAINT

ONE_CONSTRAINT_TERM -> WORDS CONSTRAINT | WORD_OR_WORDS CONSTRAINT S WORD_OR_WORDS

ONE_CONSTITUENT -> ONE_CONSTITUENT_TERM | CONSTITUENT

ONE_CONSTITUENT_TERM -> ONE_CONSTITUENT_PREFIX | WORD_OR_WORDS S CONSTITUENT

ONE_CONSTITUENT_PREFIX -> CONSTITUENT S WORD_OR_WORDS | WORD_OR_WORDS S CONSTITUENT S
WORD_OR_WORDS

#-----------------------------------

# Basic definitions

#-----------------------------------

# Define CONSTRAINT

CONSTRAINT -> "[" CONSTRAINT_TYPE ":" WORD_OR_WORDS "]"

# A CONSTITUENT only has WORDS

CONSTITUENT -> "(" WORDS ")"

# Allow MORPH to be a phrase

MORPH -> "<" WORD_OR_WORDS ">"

# Core WORD definitions

WORD_OR_WORDS -> WORD | WORDS

WORDS -> WORD S WORDS | WORD S WORD

WORD -> W

W -> C W | C

#-----------------------------------

# Terminals

#-----------------------------------

EX -> "E" "X" "A" "M" "P" "L" "E" "_" "O" "F"

CONSTRAINT_TYPE -> "e" "v" "f" | "s" "y" "n" | "h" "y" "p"

C -> "a" | "b" | "c" | "d" | "e" | "f" | "g" | "h" | "i" | "j" | "k" | "l" | "m" | "n"
| "o" | "p" | "q" | "r" | "s" | "t" | "u" | "v" | "w" | "x" | "y" | "z" | "A" | "B" |
"C" | "D" | "E" | "F" | "G" | "H" | "I" | "J" | "K" | "L" | "M" | "N" | "O" | "P" |
"Q" | "R" | "S" | "T" | "U" | "V" | "W" | "X" | "Y" | "Z" | "-" | "'" | "0" | "1" |
"2" | "3" | "4" | "5" | "6" | "7" | "8" | "9"

S -> " " S | " "
```

5 Implementing a Parser

Given a CFG definition like the one above, a simple way to parse a OpenCLIR query is to use the CFG parser from Python's NLTK toolkit as follows:

```
import nltk

mql_grammar = nltk.CFG.fromstring("""<the MQL CFG>""")
```

```
mql_parser = nltk.ChartParser(mql_grammar)
for mq in list_of_mqs:
    mq_parse_tree = mql_parser.parse_one(mq, tree_class=nltk.tree.ParentedTree)
```

This will produce a parse tree for every query.

NOTE: The above code snippet combined with the CFG specification from Section 4 will fail to parse some of the example queries from Section 2 due to extra spaces in them. For example, it will fail to parse **"EXAMPLE_OF(virus) [evf: medicine]"** but will parse **"EXAMPLE_OF(virus)[evf:medicine]"**. In all examples shown in Section 3, all extra spaces have been removed.