# Corpus Query Language (CQL) overview

## 1 Overview

- Query language for a lot of corpus managers
  - NoSketchEngine
  - Project Aranea[1]
  - UAM Corpus Tool
  - ...

## 2 Syntax

### 2.1 Basics

- `[]` denotes a segment
- Any specifics are inside: `[word="kalt"]` — looks for all the wordforms `kalt`
- `[lemma="kalt"]` — looks for all the occurrences of a lemma `kalt`
- `[tag="ADJ.*"]` — `.*` helps to find all the modifications of `ADJ` tag
- `[tag="ADJ.*"& lemma="a.*"]` — looks for all adjectives that start with a letter `a`

### 2.2 Quantification

- regex quantifiers (`*`, `?`, `+`) can be used to specify a number of segments
- `[tag="ADJ.*"]+ [tag="N.*"]` — any number of adjectives + a noun
- `[lemma="kalt"] []0,4 [tag="N.*"]` — `kalt` + noun with no more than 4 tokens between

### 2.3 Matching

- Every segment can be numbered: `1:[]`
- Global conditions may be applied: `(1:[] 2:[]) & 1.lemma = 2.lemma` — looks for two same words together

[1] `http://unesco.uniba.sk/`