

# Coveo Enterprise Search

## How the Facets and Injection Algorithm Work

This document will explain how the facets algorithm works and will outline the major differences between CES 4.0 - 5.1, CES 5.2 and CES 6.0.

The first important thing to know about the facets algorithm is that it tries to return the best values for every field but has to use heuristics in order to have a fast running time. It is possible to notice some missing values for precise queries because the algorithm evaluates the most occurring values first, and those values might not be as frequent in the result set as in the entire index. The algorithm also has to limit itself to a finite number of facets to evaluate to ensure a fast running time. This is where the concept of injection comes into play and every version of CES has a different injection algorithm.

**INJECTION:** The action of sending values from the result set to the facets algorithm to make sure uncommon values that are present in the results set get a fair chance at showing up in the final facet values.

### CES 4 to 5.1 Injection

CES 4 to CES 5.1 use the first version of the injection. This algorithm does not require any configuration, but only works for the results that are actually displayed. That is, if the query returns 10 results to display, only those results will contribute values to the facets algorithm. Also, information has to be kept at indexing time for injection.

**Pros:** Nothing to configure, fast.

**Cons:** No control on the injection process, small footprint on disk.

### CES 5.2 Injection

CES 5.2 uses a different injection algorithm. The objective being to provide more precise results for certain fields. You can specify a list of fields for which you want very thorough injection (**GroupByInjectionFields** in **PhysicalIndex** section of the configuration file). You can also specify the depth to use (**GroupByMaxToInject** in **PhysicalIndex** section of the configuration file). By default, depth is set to 30. That means that the default injection will not be done on unspecified fields but if you specify them, the injection will be done on 3 times more results than with the previous versions of CES.

**Pros:** More control on the injection for important facet fields.

**Cons:** Hard to configure as it injects only email fields by default, moderate footprint on disk.

### CES 6.0 Injection

CES 6.0 profits from the earlier approaches and provides the best injection and facets algorithm to date. That means that injection is entirely automatic without any need to configure fields for it. Also, no structure is built at indexing time for injection, thus saving disk space. Finally, each field can be configured to be scanned entirely. It is then possible to get all the possible field values, with a proportional query execution time cost.

**GroupByMaxToProcess** in **PhysicalIndex** section of the configuration file: This setting will determine how many values to scan for each field in the index. The default is 128 and is often more than enough.

This means that the 128 most frequent values will be injected. If a specific field has more than 128 values, it is possible to go deeper by increasing the value of this setting.

**GroupByFieldsWithoutLimit** in **PhysicalIndex** section of the configuration file: This setting will list the fields for which there is no limit to how many values are scanned during the facets algorithm. This will return every value if asked for them, but will have a performance hit proportional to the number of different values the fields have. Use this only when absolutely necessary.

**Pros:** Total control on the injection and execution for all fields or specific fields, no footprint on disk, does the right thing out of the box.