



# Entireweb Search API

## Query Language Specification

Entireweb Sweden AB  
Last revision: 2010-04-28

This specification documents the particulars of formulating questions to the Entireweb and ExpiCore search engines. The first section gives a formal, precise definition of the language. We then proceed to give several examples and usage tips, and also detail all special (“tagged”) words that can be used in searches.

## ***Backus-Naur Specification of Entireweb Query Language***

```
<query> ::= <optional_whitespace> <expression_list>
<whitespace> ::= [\x01-\x1F]+ <!-- All ASCII 0x01 to 0x1F are considered whitespace -->
<optional_whitespace> ::= "" | <whitespace> <optional_whitespace>
<expression_list> ::= <expression> | <expression> <query>
<expression> ::= <parenthesized_expression> |
<prefix_expression> |
<infix_expression>
<parenthesized_expression> ::= "(" <optional_whitespace> <expression> <optional_whitespace> ")"
<prefix_expression> ::= <negated_expression> | <forced_expression>
<negated_expression> ::= "-" <expression>
<forced_expression> ::= "+" <expression>
<infix_expression> ::= <entity> { <optional_whitespace> <infix_operator> <optional_whitespace> <entity> }
<infix_operator> ::= <and_operator> | <or_operator>
<and_operator> ::= "AND" | "and" | "&"
<or_operator> ::= "OR" | "or" | "|"
<entity> ::= <entity_base> | <entity_base> <weight>
<weight> ::= "$$" [0-9]+
<base_entity> ::= <term> | <phrase> | <tagged>
<phrase> ::= "" <optional_whitespace> <term> { <whitespace> <term> } <optional_whitespace> ""
<tagged> ::= <tag> ":" <term>
<tag> ::= [A-Za-z]+
<term> ::= [A-Za-z0-9\-\.\_\x80-]+
```

## ***Further Details and Examples of the Query Language***

The query language is in essence a hierarchical tree representation of entities to look up from an index, and operators describing how to process and combine these entities. An entity is essentially a word, but generalized to include multiword-phrases and special tagged constructs. The query is evaluated left-to-right using normal set-operator logic and precedence. AND is implied if no operator is provided between terms. Parentheses can be used to alter evaluation order. Now follow some examples:

- **entireweb search engine**: All three terms should exist in the document
- **entireweb OR windseek search**: Since AND has higher precedence than OR, this searches for documents containing entireweb or both the terms windseek and search.
- **(entireweb OR windseek) search**: This is probably the intended query in the last example: Searches for documents containing entireweb and search, or windseek and search.
- **-entireweb "search engine"**: Search for documents containing the precise phrase "search engine", but not the term entireweb.
- **-(entireweb OR -(windseek AND -citosearch)) "search engine"**: It may not be obvious what results this query will yield. Using de Morgan's theorem, this can be rewritten as **-entireweb AND (windseek AND -citosearch) AND "search engine"**. The commutativity and associativity of the AND operator then yields **windseek "search engine" -entireweb -citosearch**. This illustrates the fact that Entireweb operates using normal set algebra.
- **entireweb search +engine**: Sometimes, the search engine can suppress or relax terms in a search query, if they are found to contribute little to the result. Use the + (force) operator to inhibit this relaxation and force the term to be considered.
- Search terms and operator names are case-insensitive. **cAr oR bIkE** yields exactly the same results as **car or bike**.
- Whitespace may be inserted anywhere in the query, except immediately after prefix operators and before weights.
- Custom weights can be assigned to entities, to alter their importance in the query. The weights should be between 0 and 1000, and should appear immediately after the entity, preceded by two \$ signs. The weights will be renormalized. Thus the query **car\$\$1000 bike\$\$10** indicates a query in which the word car should be valued much higher than the word bike.
- Tagged entities are terms preceded by a tag. The tag specifies the precise meaning and interpretation of the entity. A tagged entity can never be negated: if it appears negated, it will be dropped from the query entirely. Also, there are special rules for handling queries with multiple tagged phrases – some tags may take precedence over others. A later section in this document describes the set of tagged entities that Entireweb understands.

## Tagged Entities

Entireweb supports multiple tagged entities, which are not directly terms retrieved from the index, but rather metadata that may alter the query processing. The following gives a list of all recognized tags. A tagged entity has the general form tag:value.

- h, http, host, site, domain: Return only documents from the host described by value. If value is preceded by a period, it retrieves documents from any subdomain that has value as a suffix, otherwise documents are returned from the particular hostname provided in value. The subdomain www is always omitted from a hostname. Thus the tagged phrase **host:.entireweb.com** returns documents from eg [www.entireweb.com](http://www.entireweb.com) and services.entireweb.com, but **host:entireweb.com** returns documents from the precise hosts entireweb.com and [www.entireweb.com](http://www.entireweb.com).
- p page: Return only documents with a particular page file name. Example: **p:contact.php**.
- u, url: Return documents with value as a leftmost prefix of the URL. Example: **url:[www.entireweb.com/newsletter/](http://www.entireweb.com/newsletter/)** would return eg [www.entireweb.com/newsletter/advertise/](http://www.entireweb.com/newsletter/advertise/).
- f, type, file, ft, ext: Return only documents of a particular file type ("extension"). Example: **f:pdf**.
- ip: Return only documents from servers having a particular IP number or class C prefix. Example: **ip:89.150.197**.
- kw: The term occurs in a META KEYWORDS tag on the page. Example: **kw:entireweb**
- au: The term occurs in a META AUTHOR tag on the page. Example: **au:fredrik au:andersson**
- ge: The term occurs in a META GENERATOR tag on the page. Example: **ge:frontpage**
- m: There is a META tag on the page whose name equals the value of the term. Useful for field searches in well-formed document collections, where meta tags are used to group documents into collections. Example: **m:rating**
- Any other tag preceding a comma: The value occurs in a meta tag with that name. Example: **rating:general**
- id: Return ranking information about a particular document having value as its internal ID number. This feature is only available to certain specific partners. See separate documentation for this entity. Example: **id:83362351045756432**.