

This guide includes an overview of the IP.com full text search facilities, including syntax, available Boolean and position operators, and functions such as wildcards and numeric ranging. You may print a copy of this paper for your personal reference.

Syntax Overview

A search request consists of a group of words or phrases linked by connectors such as *and* and *or* that indicate the relationship between them. Examples:

<code>apple and pear</code>	Both words must be present
<code>apple or pear</code>	Either word can be present
<code>apple w/5 pear</code>	<i>Apple</i> must occur within 5 words of <i>pear</i>
<code>apple not w/5 pear</code>	<i>Apple</i> must not occur within 5 words of <i>pear</i>
<code>apple and not pear</code>	Only <i>apple</i> must be present
<code>name contains smith</code>	The field <i>name</i> must contain <i>smith</i>

If you use more than one connector, you should use parentheses to indicate precisely what you want to search for. For example, `apple and pear or orange` could mean `(apple and pear) or orange`, or it could mean `apple and (pear or orange)`.

Search terms may include the following special characters:

?	Matches any single character	<code>appl?</code> matches apply or apple
*	Matches any number of characters	<code>appl*</code> matches apply, application, or apples
%	Fuzzy search	<code>ba%nana</code> matches banana, bananna
#	Phonic search ¹	<code>#smith</code> matches smith, smythe
~	Stemming ²	<code>apply~</code> matches apply, applies, applied
~~	Inclusive numeric range	<code>12~~24</code> matches 18
:	Variable term weighting	<code>apple:4 w/5 pear:1</code>

Words and Phrases

You do not need to use any special punctuation or commands to search for a phrase. Simply enter the phrase the way it ordinarily appears. You can use a phrase anywhere in a search request. Example:

`apple w/5 fruit salad`

Punctuation inside of a search word is treated as a space. Thus, *can't* would be treated as a phrase consisting of two words: *can* and *t*. *1843(c)(8)(ii)* would become *1843 c 8 ii* (four words).

Note that all words, regardless of their “semantic value,” are indexed. This includes words that are sometimes considered “noise,” such as *the*, *a*, or *of*. Therefore, searching for `the theory of relativity` is likely to return better results when expressed as `theory w/2 relativity`.

Wildcards (* and ?)

A search word can contain the wildcard characters * and ?. A ? in a word matches any single character, and a * matches any number of characters. The wildcard characters can be in any position in a word. For example:

`appl*` Matches *apple*, *application*, etc.
`*cipl*` Matches *principle*, *participle*, etc.
`appl?` Matches *apply* and *apple* but not *apples*
`ap*ed` Matches *applied*, *approved*, etc.

Use of the * wildcard character near the beginning of a word may noticeably slow searches.

Fuzzy Searching

Fuzzy searching will find a word even if it is misspelled. For example, a fuzzy search for `apple` will find *apple*. Fuzzy searching can be useful when you are searching text that may contain typographical errors or for text that has been scanned using optical character recognition (OCR). You can add fuzziness selectively using the % character. The number of % characters you add determines the number of differences the search facility will ignore when searching for a word. The position of the % characters determines how many letters at the start of the word have to match exactly. Examples:

`ba%nana` Word must begin with *ba* and have at most one difference between it and *banana*.
`b%anana` Word must begin with *b* and have at most two differences between it and *banana*.

Note that a significant number of the documents submitted to the IP.com Prior Art Database include text created using OCR technologies.

Phonic Searching

Phonic searching¹ looks for a word that sounds like the word you are searching for and begins with the same letter. For example, a search for `#smith` will find *Smithe* and *Smythe* as well as *Smith*; a search for `#fiber` will find *fiber*, *fibre*, and *fibers*.

To ask the search facility to search for a word phonically, place a # in front of the word in your search request.

Stemming

Stemming² extends a search to cover grammatical variations on a word. For example, a search for `fish~` would also find *fishing*. A search for `apply~` would also find *applying*, *applies*, *applied*, as well as *apply*.

To use stemming, add a ~ at the end of words that you want stemmed in a search.

Numeric Range Searching

A numeric range search is a search for any numbers that fall within a range. To add a numeric range component to a search request, enter the upper and lower bounds of the search separated by `~~` like this:

```
apple w/5 12~~17
```

This request would find any document containing *apple* within 5 words of a number between *12* and *17*. Numeric range searches only work with positive integers. A numeric range search includes the upper and lower bounds (so *12* and *17* would be retrieved in the above example).

For purposes of numeric range searching, decimal points and commas are treated as spaces and minus signs are ignored. For example, `-123,456.78` would be interpreted as: *123 456 78* (three numbers).

Variable Term Weighting

When the search facility sorts search results after a search, by default all words in a request count equally in counting hits. However, you can change this by specifying the relative weights for each term in your search request, like this:

```
apple:5 and pear:1
```

This request would retrieve the same documents as *apple* and *pear* but, the search facility would weight *apple* five times as heavily as *pear* when sorting the results.

Field Searching

When the search facility indexes packages or publications, it saves certain information so that you can perform searches limited to a particular field. For example, suppose the database has a *NAME* field and a *DESCRIPTION* field. You could search for *apple* in the *NAME* field like this:

```
NAME contains apple
```

Field searches can be combined using *and*, *or*, and *not*, like this:

```
(CITY contains (boston or york)) and (ADDRESS contains (washington))
```

The parentheses are necessary to ensure that the search facility interprets the search request correctly.

Special IP.com InnovationQ Field Names

IP.com's InnovationQ packages contain certain information that may be accessed using field searching. The built-in fields are as follows:

TITLE	The provided title or the name of the required file.
ABSTRACT	The provided abstract/keywords or an automatically extracted leading portion of the required file.
FILENAME	All the filenames included in all the packages.
RELATED_PERSON	The screen name and email address of the submitter (at the time of submission).

In addition to these built-in fields, an InnovationQ collection may contain additional named fields. If any are available, their names are indicated at the end of the "...matching these custom fields" section of the search screen. Custom field names always begin with the dollar (\$) symbol.

Special IP.com Prior Art Database Field Names

The IP.com Prior Art Database contains some information that may be utilized using field searching. Note that the inclusion of these fields and the quality of the information they contain is decided by the submitter of the document, not by IP.com. The fields are as follows:

TITLE Generally, submitters provide descriptive titles for documents.

ABSTRACT Most, but not all, documents have useful abstracts.

COPYRIGHT Some documents have copyright information; the format of the information is unpredictable.

CLASSIFICATION Few documents have classification information; the format is either WIPO's IPC7 system or the USPTO's system.

RELATED_PEOPLE This field includes all information regarding each related person. Some documents contain related people information.

RELATED_DOCUMENTS Includes all information regarding each related document. Few documents contain related document information.

ORIGINAL_SOURCE Certain documents are members of a specific collection of information; if present in a document, this field contains an indication of this collection. See below.

ORIGINAL_DATE For those documents that contain an **ORIGINAL_SOURCE**, the original date of the publication. Depending on the nature of the information, this may not be the same as the IP.com publication date.

IP.com Prior Art Database Corporate Disclosure Subsets

The IP.com Prior Art Database contains the following collections of historic and current documents. Use the **ORIGINAL_SOURCE** field to distinguish these documents:

ORIGINAL_SOURCE contains

IBM-TDB IBM Technical Disclosure Bulletins not originating in the IP.com Prior Art Database.

IBM-IPCOM IBM Technical Disclosures originally published in the IP.com Prior Art Database.

MOT-TDB The Motorola Technical Bulletins not originating in the IP.com Prior Art Database. Available from August 1980.

MOT-IPCOM Motorola Technical Bulletins originally published in the IP.com Prior Art Database.

SI-TDB The Siemens Technical Bulletin not originating in the IP.com Prior Art Database.

SI-IPCOM Siemens Technical Bulletin originally published in the IP.com Prior Art Database.

The **ORIGINAL_DATE** field contains month and year information in the form *YYYYMM* (*YYYY* is the year, *MM* is the month). This is the month and year the document originally appeared in print if it was not originally published using the Prior Art Database or it repeats the Prior Art Database month and year of publication. The format allows the use of the *~~* numeric range operator.

For example, to select all 1999 IBM publications, use the query:

```
(original_source contains IBM*) and  $\hat{E}$ 
(original_date contains 199901~~199912)
```

The Internet Society's Requests for Comments

As a service to the International Patent Offices, the IP.com Prior Art Database includes the Internet Society's Requests for Comments (the ISOC's RFCs). These documents are available as free downloads for all users and are updated regularly from the data made available at <ftp://ftp.rfc-editor.org/in-notes>.

More information about the Internet Society is available at <http://www.isoc.org>.

To distinguish these documents use `ORIGINAL_SOURCE` contains ISOC-RFC. The `ORIGINAL_DATE` field is also available for this subset.

Boolean and Positional Connectors

The AND Connector

Use the `AND` connector in a search request to connect two expressions, both of which must be found in any document retrieved. For example:

<code>apple pie and poached pear</code>	Retrieve any document that contains both phrases
<code>(apple or banana) and (pear w/5 grape)</code>	Retrieves any document that (1) contains either <i>apple</i> or <i>banana</i> and (2) contains <i>pear</i> within 5 words of <i>grape</i>

The OR Connector

Use the `OR` connector in a search request to connect two expressions, at least one of which must be found in any document retrieved. For example:

<code>apple pie or poached pear</code>	Retrieves any document that contains <i>apple pie</i> , <i>poached pear</i> , or both
--	---

The W/# Connector

Use the `W/#` connector in a search request to specify that one word or phrase must occur within some number of words of the other. For example:

<code>apple w/5 pear</code>	Retrieves any document that contains <i>apple</i> within 5 words of <i>pear</i>
-----------------------------	---

The following are examples of search requests using `W/#`:

```
(apple or pear) w/5 banana
(apple w/5 banana) w/10 pear
(apple and banana) w/10 pear
```

Some types of complex expressions using the `W/#` connector will produce ambiguous results and should not be used. The following are examples of ambiguous search requests:

```
(apple and banana) w/10 (pear and grape)
(apple w/10 banana) w/10 (pear and grape)
```

In general, at least one of the two expressions connected by `W/#` must be a single word or phrase or a group of words and phrases connected by `OR`. Examples:

```
(apple and banana) w/10 (pear or grape)
(apple and banana) w/10 orange tree
```

The NOT and NOT W/# Connectors

You may use **NOT** in front of any search expression to reverse its meaning. This allows you to exclude documents from a search. Example:

```
apple sauce and not pear
```

NOT standing alone can be the start of a search request. For example, the following would retrieve all documents that did not contain *pear*:

```
not pear
```

If **NOT** is not the first connector in a request, you need to use either **AND** or **OR** with **NOT**:

```
apple or not pear  
not (apple w/5 pear)
```

The **NOT W/** ("not within") operator allows you to search for a word or phrase not in association with another word or phrase. Example:

```
apple not w/20 pear
```

Unlike the **W/** operator, **NOT W/** is not symmetrical. That is, `apple not w/20 pear` is not the same as `pear not w/20 apple`. In the `apple not w/20 pear` request, the facility searches for *apple* and excludes cases where *apple* is too close to *pear*. In the `pear not w/20 apple` request, the facility searches for *pear* and excludes cases where *pear* is too close to *apple*.

Non-English Language Searching

As noted earlier, the search facility natively operates in English when considering phonic and stemming rules. However, searching for any language is fully enabled. Simply enter the non-English words (containing the appropriate characters) as you normally would. It is therefore possible to search for a German-titled document like this:

```
title contains (ortsabhängige w/5 teilnehmerverfügbarkeit)
```

To fully utilize the capabilities of the search facilities, note that the individual characters in all Eastern language (Chinese, Korean, and Japanese) documents are indexed as individual words. In Japanese, this is true regardless of whether the actual word is a combination of Kata and Kana characters.

Credits

This document was adopted from the [dtSearch Version 6 User's Manual](#), Chapter 6 "Search Requests." dtSearch's Text Retrieval Engine powers IP.com's online full-text search facility.

Visit dtSearch Corp at <http://www.dtsearch.com>.

Visit IP.com at <http://www.ip.com>.

¹ Only English language phonetic rules are applied.

² Stemming rules are limited to the English language; only English matches will be reliably located.