INTERNATIONAL STANDARD

ISO
8777

Information and documentation — Commands for interactive text searching

Information et documentation — Commandes pour les systèmes interactifs de recherche d'information

Contents

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

International Standard ISO 8777 was prepared by Technical Committee ISO/TC 46, *Information and documentation*, Sub-Committee SC 4, *Computer applications in information and documentation*.

Annexes A, B, C and D of this International Standard are for information only.

Introduction

A consequence of the many information retrieval systems that exist today is the great number of different commands, types of command language and even search philosophies in existence. A way to facilitate the use of such systems, particularly for the more infrequent user, is to create a standard user interface consisting of a basic set of commands and rules for their use. This International Standard is intended for that purpose.

Many types of user-system interface are used with interactive search systems, and this International Standard, recognizing this variety, is addressed to one particular type of interface, namely, command-driven systems.

The mode of operation of search systems complying with this International Standard is often known as user-driven dialogue, where most actions are initiated by the user issuing commands to this system; however, this does not preclude the system requiring answers to specific questions on occasion.

Information and documentation — Commands for interactive text searching

1 Scope

This International Standard specifies a basic set of commands for the interactive search of retrieval systems data and the types of response expected from the processing system. It is intended for use by designers and users of information retrieval systems, including computer-based library catalogues and computer-based database access and search facilities.

This International Standard does not restrict or prohibit the use of other types of user-system interaction, such as menus or natural language interfaces, or the use of a "native" (non-standard) command language.

NOTE 1   If the meaning of a standard command name conflicts with the meaning of a command name in the "native" language, the function of the standard command name is preferred.

Some systems incorporate fewer functions than specified in this International Standard; others use additional functions. These additional functions may include added command names, operators or qualification and limitation techniques. It is important that additional functions and the type of response they generate conform to the general features and syntax set out in this International Standard.

2 Normative reference

The following standard contains provisions which, through reference in this text, constitute provisions of this International Standard. At the time of publication, the edition indicated was valid. All standards are subject to revision, and parties to agreements based on this International Standard are encouraged to investigate the possibility of applying the most recent edition of the standard indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

ISO/IEC 646:1991, *Information technology — ISO 7-bit coded character set for information interchange.*

3 Definitions

The definitions in this International Standard are intended to reflect the point of view of the user. They are not intended to reflect the technical or engineering aspects of computers. For the purposes of this International Standard, the following definitions apply.

3.1 basic index: Index associated with *fields* that are searched when no *field* designations are stated.

3.2 command expression: Complete request for the performance of a function.

3.3 command name: Specially defined *reserved word* or abbreviation used to initiate a *command expression*.

3.4 command specification: String of characters following a *command name*, specifying how and on what the *command expression* is to operate.

3.5 connector: Symbol used to link *search terms* and *qualifiers*.

3.6 default: Value automatically assumed by the system, unless the user specifies a different value or values.

3.7 field: As a subset of a *record*, structured data treated as a unit and used to store a particular type of data.

3.8 field label: String of characters used to uniquely identify a particular *field* defined in a *record*.

3.9 masking: Symbolizing unknown or unspecified character(s) in a *search term* by special characters that are defined to represent any character or characters, no characters, or blank spaces.

3.10 operator: *Reserved word* or symbol used to specify the relationship between two entities being searched. Operators include the following.

3.10.1 Boolean operator: *Operator* that expresses logical relationships between two *search terms* or *search elements*, e.g. AND, NOT and OR.

3.10.2 proximity operator: *Operator* that specifies the sequence of and distance between two *search terms*.

3.10.3 ranging operator: *Operator* that assigns ranges of consecutive values between two *search terms*.

3.11 parameter: Variable that is given a constant value for a specified application, and that may denote that application. [ISO 2382-2]

3.12 qualifier: Parameter used to restrict or otherwise direct the range of values of a given variable.

3.13 record: Group of data usually comprising a number of *fields* and treated as a unit.

3.14 reserved word: *Word*, abbreviation or symbol that has a special meaning defined explicitly in the command language.

3.15 restoration mark: Symbol used to restore the literal meaning of a defined *reserved word*.

3.16 result set: Group of *records* retrieved by a *search statement*.

3.17 result set identifier: Label assigned by the system or by the user to a *result set*. The result set identifier for a given *search statement* is identical to its *search statement identifier*.

3.18 (online) search: Interactive search process via computer to as many databases as the searcher considers necessary to find the requested information.

3.19 search element: *Search term*, or Boolean combination of *search terms* to be searched in the same index, and its *qualifier* or *qualifiers* (the *qualifier* may be implicit), or a *result set identifier*, or terms selected from the display resulting from a SCAN or RELATE command expression.

3.20 search statement: *Command specification* of a particular FIND command.

3.21 search statement identifier: Label assigned by the system to each *search statement*.

3.22 search strategy: Series of command expressions intended to satisfy a request for information. A search strategy may include a variety of

database selection commands, term identification commands, and search and display commands.

3.23 search term: *Word* or group of *words* that the FIND command instructs the system to retrieve. A search term contains *search words* and may contain *proximity operators* but does not include *Boolean operators* or *ranging operators* or *qualifiers*.

3.24 search word: *Word* that the system is capable of searching.

3.25 separator: Character or series of characters used to set apart components of a command expression. Space, colon, semicolon and parentheses are defined in this International Standard as separators.

3.26 session: All transactions between user and system from logon to logoff.

3.27 stopword: *Word* ignored by the indexing of a particular database.

3.28 thesaurus: Compilation of terms showing synonymous, hierarchical and other relationships and dependencies, the function of which is to provide a previously-built controlled vocabulary for information storage and retrieval.

3.29 truncation: Special form of character *masking* that masks a character or characters at either end of a word.

3.30 word: Character or characters preceded and followed by *separators*. The characters may be alphanumeric or symbols.

4 General principles

4.1 Implementation

This International Standard does not specify the manner in which system designers implement the functions it describes. These command functions and the expected system responses are described from the user's point of view.

4.2 Conformance

An information retrieval system conforms to this International Standard when it recognizes and responds to every command specified by this International Standard. Where a facility is unavailable, the system response shall inform the user of this.

4.3 Command structure

In this International Standard, the following general command structure is applied:

< command expression > =
< command name > < command specification >

A command expression shall start with a command name or abbreviation. Not all command expressions require a command specification.

4.4 Command names

4.4.1 General

Command names are considered international in language and are selected to describe the function to be performed. See table 1 for a summary of command names.

The following criteria have been followed in the selection of command names and shall be followed in the formulation of additional command names:

a) the number of command names should be kept to a usable minimum;

b) verb forms are preferred;

c) command names should be as self-explanatory as possible.

4.4.2 Command name abbreviations

Command names shall be abbreviated by truncation of characters from the right. For all command names in this International Standard, the first three characters are specified as the standard abbreviation (see annex A). The system is required to accept both the full form of the command name and the three-letter abbreviation.

In addition, any unambiguous truncated form of the command name ranging from the initial character to the full command name shall be recognized by the system. If ambiguity results from the user's abbreviation in any case, the system should respond by asking for a fuller, unambiguous form of the command name.

4.5 Command specification components and format

4.5.1 Components

A command specification may contain user-supplied data such as:

a) search terms;

b) system-defined qualifiers (e.g. field labels);

NOTE 2   Commonly used field labels are given in annex B.

c) record or format identifiers;

d) term identifiers (e.g. those resulting from a SCAN or RELATE command expression);

e) standard-defined Boolean, proximity and ranging operators;

f) character masking symbols;

g) result set identifiers.

If the command specification lacks the necessary details, the system shall respond either by asking the user to supply essential information or according to preset default values.

4.5.2 Format

The order and format of the components in a command expression are specified in this International Standard.

4.6 Character codes

This International Standard is designed to be used with the 7-bit code specified in ISO/IEC 646, and all references to characters or special symbols refer to that code. Systems using other codes capable of providing the same or equivalent graphic characters may be used.

4.7 Characters

The system shall be capable of accepting user input without discriminating between upper and lower case.

4.8 Separators

4.8.1 Spaces

Spaces are significant and are used to separate components of a command expression. A space shall appear after a command name when it is followed by a command specification. A system conforming to this International Standard shall accept multiple spaces as a separator equivalent to a single space.

In addition, a space between search words shall be interpreted as a proximity operator requiring immediate adjacency of the two words in the order entered.

4.8.2 Commas

The comma (,) shall be used as a special purpose separator to separate two or more like command components, such as qualifiers (e.g. field labels) or record identifiers (e.g. record numbers). The system shall accept any number and any combination of spaces and commas together as a single comma.

The treatment of commas that occur within the text of data fields, such as within author fields that use inverted names, is implementation-specific and is not dictated by this International Standard. In any other cases where the presence of a comma causes conflict with the rules of this International Standard, restoration marks shall be used to preserve the textual nature of the comma.

4.8.3 Semicolons

The semicolon (;) shall be used to separate command expressions in user-defined sequences of command expressions, or for command "stacking" purposes, that is, when several commands are submitted to the system in a single transmission.

4.8.4 Parentheses

Parentheses act as implicit separators; that is, a left parenthesis shall not require a preceding space, nor shall a right parenthesis require a following space. A space immediately inside a parenthesis shall be ignored.

Parentheses shall be used to specify precise groupings of elements in a search statement to ensure that a sequence of operators is executed in the intended order.

5 Command names

The command names specified in table 1 shall be used, in accordance with clauses 6 to 17, in constructing the command expressions specified in this International Standard.

6 User aids

6.1 System guidance: INFO

The command name INFO shall be used to obtain information about the system, its database(s), or other features. This information is the same at any point in the session.

INFO does not require a command specification but may be used with one. When INFO is entered alone, the system shall display the list of available topics.

Only one topic shall be specified in the command specification. If the topic specified by the user is not available, the system shall display a list of available topics.

See C.1.1 for examples of the use of INFO.

6.2 Session guidance: HELP

The command name HELP shall be used to obtain online assistance or instruction specific to the user's situation or to the context of the interaction.

HELP does not require a command specification. No particular command specification is defined in this International Standard.

See C.1.2 for examples of the use of HELP.

Table 1 — Basic command names

Principal application	Command name	Relevant clause or subclause for command
User aids	INFO	6.1
	HELP	6.2
	REVIEW	6.3
Paging	FORWARD	7.1
	BACK	7.2
Database selection	BASE	8
Query formulation	FIND	9
Index scanning	SCAN	10
Thesaurus scanning	RELATE	11
Output commands	SHOW	12
	PRINT	12
Saving and recalling search strategies	SAVE	13
Deletion	DELETE	14
User-defined functions	DEFINE	15
Interrupt	not specified	16
End of session	STOP	17

6.3 Search history: REVIEW

The command name REVIEW is used to review the search elements and search statements entered during the session and still available for use.

REVIEW does not require a command specification but may be used with one. When REVIEW is used alone, the system shall provide a list of all search statements entered during the session including the number of retrieved records for each statement and the search statement or result set identifiers.

The following command specifications for REVIEW shall be available:

a) search statement identifier(s); to review the statements identified including the number of retrieved records. The identifiers may be given as a range $< n >$ TO $< m >$.

b) SAVE=$<$ saved search strategy identifier $>$; to recall a saved search strategy without reprocessing.

See C.1.3 for examples of the use of REVIEW.

6.4 Numbering: $<$ number $>$

Numbers or other means of identification shall be used to identify:

a) search statements;

b) terms displayed in response to a SCAN or RELATE command expression;

c) records retrieved in response to a search;

d) print command expressions; and

e) saved searches.

While the precise form of identifiers may differ from system to system, each series shall be separately labelled and, if numeric, consecutively listed.

7 Paging

7.1 FORWARD

FORWARD shall be used in any application to view continuing data, or data following displayed data or items on a list.

FORWARD does not require a command specification but may be used with one. If no command specification is given, the system default shall apply.

The following command specifications shall be available for the FORWARD command expression:

a) $< n >$, a positive integer; to move forward n screens or pages;

b) REC $< n >$, where n is a positive integer; to move forward n records, regardless of the number of screens or pages those records occupy.

FORWARD is not a substitute for a search statement or other command expression. A FORWARD command expression will usually be entered after a command expression initiated by DISPLAY, RELATE, REVIEW, or SCAN has been executed, as well as after any other command expression whose response fills more than one screen or page.

See C.2.1 for examples of the use of FORWARD.

7.2 BACK

BACK shall be used in any application to view preceding data, or data coming before displayed data or items on a list.

BACK does not require a command specification but may be used with one. If no command specification is given, the system default shall apply.

The following command specifications shall be available for the BACK command expression:

a) $< n >$, a positive integer; to move backward n screens or pages;

b) REC $< n >$, where n is a positive integer; to move backward n records, regardless of the number of screens or pages those records occupy.

BACK is not a substitute for a search statement or other command expression. A BACK command expression will usually be entered after a command expression initiated by DISPLAY, RELATE, REVIEW, or SCAN has been executed, as well as after any other command expression whose response fills more than one screen or page.

See C.2.2 for examples of the use of BACK.

8 Database selection: BASE

The command name BASE shall be used to select the database(s) to be searched.

BASE does not require a command specification but may be used with one. When BASE is entered alone, the system shall respond with a list of available databases and instructions on how to select one or more.

When BASE is used with a command specification, the system shall confirm the name(s) for the database(s) and give the time span(s) for the databse(s).

Database names are system-dependent.

If simultaneous access to more than one database is possible, the names of the databases in the command

specification shall be separated by commas. A name assigned to a group of databases is acceptable.

It shall be possible to use masking symbols in database names.

See C.3 for examples of the use of BASE.

9 Query formulation: FIND

9.1 General

The command name FIND shall be used for query formulation. It invokes a search of one or more of the system's databases.

9.2 Search statements

The command specification for FIND is mandatory and is known as the search statement.

A search statement consists of a search element or a combination of search elements.

< search statement >=< search element >< operator >< search element >

A search element may consist of:

a) a search term, or a Boolean combination of search terms to be searched in the same index, and its qualifier or qualifiers (the qualifier may be implicit); or

b) a ranging expression with its qualifiers; or

c) a result set identifier; or

d) terms identified through a SCAN or RELATE command expression.

A search term contains search words and may contain proximity operators but shall not contain Boolean or ranging operators, or qualifiers.

Example of the components of a FIND command expression

Command expression

FIND s4 AND (mark ! twain OR samuel ! clemens) AND TI,SU tom sawyer AND DA LT 1900

Command name

FIND

Search words

mark
twain
samuel
clemens
tom
sawyer
1900

Search terms

mark ! twain
samuel ! clemens
tom sawyer
1900

Search elements

s4
(mark ! twain OR samuel ! clemens)
TI,SU tom sawyer
DA LT 1900

Command specification or search statement

s4 AND (mark ! twain OR samuel ! clemens) AND TI,SU tom sawyer AND DA LT 1900

See C.4.1 for further examples.

9.3 Restoration marks

Double quotation marks (") shall enclose a reserved command name, abbreviation, operator or symbol to restore its literal meaning so that it may be used as a search term.

See C.4.2 for examples of the use of restoration marks.

9.4 Character masking

Two symbols are defined to enable the user to perform character masking and truncation. In each case, the symbol shall be embedded without intervening spaces in the word at the point of the mask. Multiple occurrences of the same symbol shall have no intervening spaces between occurrences. Different masking symbols may be used at different points within a word.

9.4.1 Masking a precise number of characters

The symbol # shall be used to indicate that precisely one character is to be masked.

Multiple #s shall be used to indicate that a precise number of characters greater than one and equal to the number of # symbols are to be masked.

See C.4.3.1 for examples of use.

9.4.2 Masking a variable number of characters

The symbol ? shall be used to indicate a variable number of characters to be masked. A single ? shall be used to indicate an unlimited number of characters. ?n, where n is a positive integer, shall be used to indicate a limited range of characters to be masked, from zero up to and including the specified integer.

See C.4.3.2 for examples of use.

9.4.3 Response

The response to a request containing masking symbols shall be either:

a) a set resulting from a combination of all terms matching the mask, using the Boolean operator OR; or

b) a list of those terms matching the mask.

If a system's default response is response b), the specification ALL, preceding the masked word, shall be used to request response a).

9.5 Operators

9.5.1 Boolean operators

The logic operators AND, NOT and OR are used to connect search elements, result sets, or sets retrieved by a RELATE or SCAN command/expression or any pair thereof.

Parentheses shall be used to ensure that a sequence of operators is executed in the intended order. The logic within the innermost matched pair of parentheses shall be executed first. Search elements may be nested, i.e. search elements may themselves contain operators, or they may be earlier search statement identifiers or result set identifiers.

Processing of Boolean operators shall occur on logical groups from left to right.

See C.4.4.1 for examples of the use of Boolean operators.

9.5.2 Ranging operators

The operators > (or GT), < (or LT), = (or EQ), < > (or NE), >= (or GE), <= (or LE) and − (or TO) shall be used to assign ranges of values to parameters in a search element. The letter equivalents shall be entered within spaces in a search element. Spaces entered on either side of the symbols shall be ignored.

The operator − (or TO) is inclusive of the start and end values; it may take nil (i.e. nothing entered) as one of its connected values.

See C.4.4.2 for examples of the use of ranging operators.

9.5.3 Proximity operators

9.5.3.1 General

Promixity operators are used to specify the relative position and distance between two search words or terms.

Three proximity operators are specified in this International Standard.

Proximity operators shall be processed on logical units from left to right.

Symbols for the proximity operators shall be separated from search words or terms by a space.

The treatment of stopwords or "common" words is not specified in this International Standard.

9.5.3.2 Direct sequence

Direct sequence means that the words shall be immediately adjacent and in the order entered. The operator for this condition shall be a single space. Multiple spaces shall be interpreted as a single space.

See C.4.4.3.1 for examples of the use of proximity operators in direct sequence.

9.5.3.3 Word order specified

The proximity operator ! shall be used between two search words in a search statement to specify precise word order. ! alone between two words shall indicate immediate adjacency in the order entered. !n, where n is a positive integer, shall be used to specify the maximum distance (number of words), n, within which the target words must occur in the order specified in the search statement.

See C.4.4.3.2 for examples of the use of proximity operators with word order specified.

9.5.3.4 Word order not specified

The proximity operator % shall be used between two search words when word order is not to be specified. % alone between two words shall indicate immediate adjacency of the words in either order. %n, where n is a positive integer, shall be used to specify the maximum number of words, n, within which the target words must occur in either order.

See C.4.4.3.3 for the examples of the use of proximity operators with word order not specified.

9.5.4 Order of precedence for evaluating operators

Operators shall be evaluated in the following order of precedence: