



Leading healthcare  
terminology, worldwide

## SNOMED CT Search and Data Entry Guide

Latest web browsable version: <http://snomed.org/searchguide>

SNOMED CT document library: <http://snomed.org/doc>

Publication date: 2017-11-22

# Table of Contents

<b>1. Introduction.....</b>	<b>4</b>
<b>2. Introduction to Search.....</b>	<b>7</b>
2.1. The Importance of Effective Search .....	7
2.2. Using SNOMED CT Features to Support Optimized Searches.....	8
<b>3. Use Cases for Searches .....</b>	<b>11</b>
3.1. Use Cases Directly Connected to Data Entry .....	11
3.2. Use Cases Where Search Browsers are Required .....	11
<b>4. Optimizing Searches .....</b>	<b>13</b>
4.1. Search by Text .....	14
4.2. Search by Identifiers .....	19
4.3. Extended Searches.....	20
4.4. Constrained Searches .....	25
4.5. Improve Search Speed .....	30
<b>5. Optimize Display of Search Results.....</b>	<b>32</b>
5.1. Order Search Results Rationally.....	32
5.2. Distinguish Identical Terms of Different Concepts.....	37
5.3. Avoid Multiple Hits on the Same Concept .....	38
5.4. Rationalize Search Results by Subsumption Checking.....	39
5.5. Display Navigation Results Effectively .....	41
5.6. Use Mnemonics and Personal Favorites for Data Entry .....	44
<b>6. Data Entry .....</b>	<b>45</b>
6.1. Benefits of Using SNOMED CT for Data Entry .....	45
6.2. Structured Data Entry .....	48
6.3. Constraining Data Entry.....	55
6.4. Entering Refinements for Post-Coordinated Expressions .....	58
<b>Appendix.....</b>	<b>62</b>
Appendix A - Using Search Index Files in the SNOMED CT Toolkit .....	62
Appendix B - Future Additions to this Guide.....	65
Appendix C - An Example Search and Data Entry Overview Table .....	65

The logo for SNOMED International, featuring the word "SNOMED" in a large, bold, white sans-serif font above the word "International" in a smaller, white sans-serif font, both set against a solid blue square background.

Leading healthcare  
terminology, worldwide

The Search and Data Entry Guide provides advice on two related activities that are essential for use of any code system: finding a term and saving the term and related code into the record. The first part of this guide is concerned with searching the content of SNOMED CT to find concepts that represent particular clinical ideas. The second part of the guide is concerned with ways to use SNOMED CT to support entry of relevant clinical information in electronic health records.

Web browsable version: <http://snomed.org/searchguide>

SNOMED CT Document Library: <http://snomed.org/doc>

© Copyright 2017 International Health Terminology Standards Development Organisation, all rights reserved.

This document is a publication of International Health Terminology Standards Development Organisation, trading as SNOMED International. SNOMED International owns and maintains SNOMED CT®.

Any modification of this document (including without limitation the removal or modification of this notice) is prohibited without the express written permission of SNOMED International. This document may be subject to updates. Always use the latest version of this document published by SNOMED International. This can be viewed online and downloaded by following the links on the front page or cover of this document.

SNOMED®, SNOMED CT® and IHTSDO® are registered trademarks of International Health Terminology Standards Development Organisation. SNOMED CT® licensing information is available at <http://snomed.org/licensing>. For more information about SNOMED International and SNOMED International Membership, please refer to <http://www.snomed.org> or contact us at [info@snomed.org](mailto:info@snomed.org).

# 1. Introduction

## Purpose of This Guide

This guide provides advice on two related activities that are essential for use of any code system: finding a term and saving the term and related code into the record. The first part of this guide is concerned with searching the content of SNOMED CT to find concepts that represent particular clinical ideas. The second part of the guide is concerned with ways to use SNOMED CT to support entry of relevant clinical information in electronic health records.

## Who Should Read This Guide

This guide should be read by anyone involved in search and data entry. These people fall into two broad categories:

### End users of applications that support SNOMED CT-enabled search and data entry

*These people need:*

- To be aware of the importance of search features that facilitate efficient and accurate SNOMED CT clinical data entry.
- To understand their own data entry requirements in order to select and use configurable search features that meet their needs.

*This guide provides:*

- A high-level description of the importance of different search and data entry features
- Examples of each of the features described.

### Those designing, developing, procuring or evaluating services that include support for SNOMED CT-enabled search and data entry

*These people need:*

- To understand the range of search features likely to be important for their users.
- To be able to develop and deploy search and data entry techniques that provide their users with effective ways to use SNOMED CT as part of the process of clinical record keeping.

*This guide provides:*

- Descriptions of search and data entry features which enable successful implementation.
- Summaries of tried and tested techniques for delivering many of these features.

## What Is Special About Searching SNOMED CT?

At face value it may seem unusual to have a specific guide for searching one terminology. General purpose search routines may seem to be applicable to any list of terms and, by comparison with internet searches, even a terminology with over a million terms does not seem too challenging. However, there are aspects of SNOMED CT that, while presenting challenges to simple searches, also enable enhancements that can deliver highly effective search optimization.

The size, design and broad scope of SNOMED CT present challenges to simple searches:

- A large terminology related to healthcare is sure to have lots of terms containing similar words (e.g. 'hip' is in more than 1,200 terms).
  - What order should these matches be shown in?
- If two or more words or parts of words are typed what should result be
  - Which should of these sets should be shown?
- The same clinical idea can be expressed in different ways
  - How can it be made easy to find an idea expressed in a different way

The logical design of SNOMED CT allows capture of meaningful clinical information and can also be used to support effective search and data entry strategies. If you see SNOMED CT as simply a long list of terms, then you are probably searching the terminology with a general purpose search tool that has overlooked the opportunities for search refinements offered by the design of the terminology.

This guide presents practical ways to make terminology searches which lead to accurate and complete clinical noting. While searches need to be fast, response times of a second or two may be acceptable if the displayed results are more relevant. A well-designed search solution should minimize the overall time to take to record clinical information. This guide does not suggest a single approach but identifies ways to address different requirements with appropriate solutions.

## Search Configuration

Many of the search techniques described in this guide can be applied to a range of different requirements. However, some search techniques require a degree of configuration to meet the needs of particular users and particular types of data collection.

Various levels of configuration should be considered including:

- Default configurations for an organization or group of users. For example, specifying the SNOMED CT version, exclusion of inactive concepts and inclusion of various extensions and supporting derivatives;
- Settings bound to data input fields that make it easy to enter commonly used values and prevent entry of inappropriate values. For example, limiting entries to a particular subset of SNOMED CT content and prioritizing some members of that set.
- User configuration that enables modification while using a front-end application to refine searches. For example, setting the search mode to include complete, partial or pattern based phrase matches or to limit searches to a particular hierarchy.

## About Search and Data Entry

Search and data entry are closely related and as illustrated in Figure 1 can be regarded as two steps in a single process. A clinical user needs to record a clinical idea (e.g. a sign, symptom, diagnosis or procedure). The user searches for the relevant Concept, views the results, finds the appropriate Concept, selects it for data entry and the data is stored.

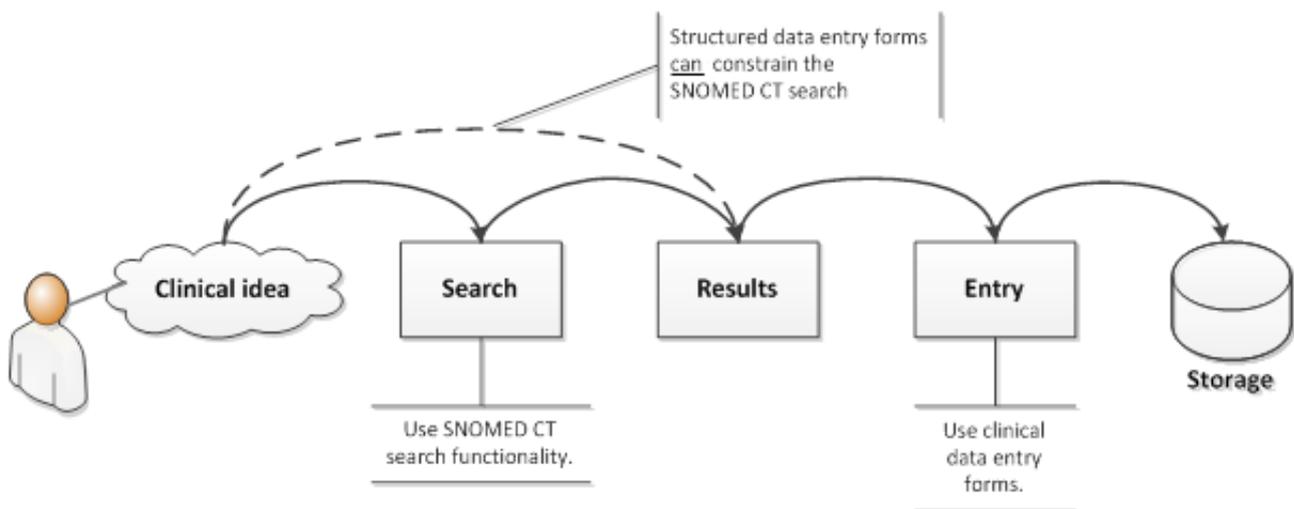


Figure 1-1: A search and data entry process

Although there is often a close relationship between search and data entry, the two processes also need to be considered separately. This guide has therefore been categorized into two parts: the Search part and the Data Entry part.

SNOMED CT searches are also required for purposes which are not directly connected to data entry including:

- Reviewing terminology content;
- Creating Reference Sets (e.g. to represent subsets of terms and Concepts);
- Creating templates and protocols for data entry;
- Creating queries to retrieve data;
- Learning about the structure of the terminology;

Please refer to the section "Use cases for searches" for further information.

Users may enter data including SNOMED CT terms using techniques that do not involve consciously searching. These include:

- Structured data entry screens with option controls linked to an appropriate SNOMED CT Concepts which are recorded based on the options selected.
- Natural language processing (NLP) including tagging of particular phrases with particular Concept identifiers. Please note that NLP techniques are out of scope of this guide.

For further information about structured data entry techniques, please refer to the "Data Entry" section.

## 2. Introduction to Search

Search is the process by which a user finds a Concept or Description to represent a clinical idea for a specific purpose.

The way a search is carried out depends on the setting in which it is performed. A simple search may involve typing a word or phrase in a search box, getting a list of matching terms and viewing the list to identify the appropriate Concept. This Guide considers a range of enhancements to SNOMED CT searches which can make it quicker and easier for users to find the appropriate Concept.

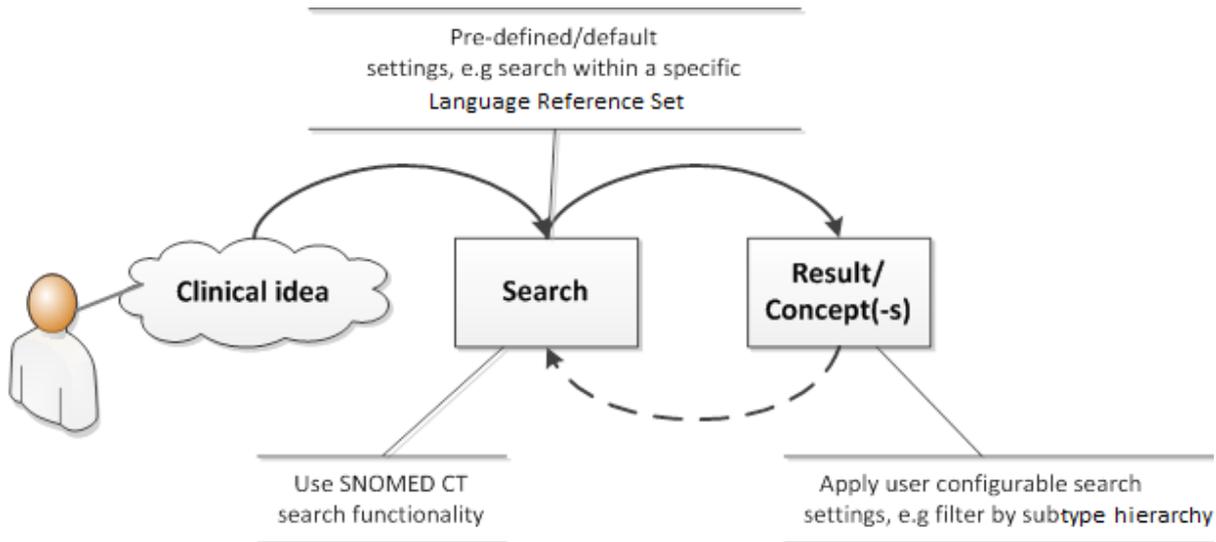


Figure 2-1: Search process

### 2.1. The Importance of Effective Search

Search is an important part of the clinical information life-cycle. Effective search must make it quick and easy for users to accurately select the relevant content for data entry. The selected Concept or Description may need to be stored in ways that represent information consistently and allow it to be processed effectively.

An EHR system must also have a user interface that makes search easy in terms of the time and effort required to find suitable Concepts to enter into the clinical record. Improving the usability of search can help increase the utility of SNOMED CT for data entry, and hence, the overall adoption of clinical systems. Increasing the utility and the added value of clinical systems depends on improving the tradeoff between time taken to code, and the quality of the coded content produced. This is true *whichever* terminology is chosen to be implemented 'front of house'.

Leveraging SNOMED CT advanced design and implementing it properly is specifically intended to give better leverage in this tradeoff than we could ever have on other coding systems. However, if badly implemented, SNOMED CT will make an already bad situation worse. On the other hand, a multitude of different but generally poor designs is equally dangerous. There is a minimum level of functionality that should be universal, and implementers need to be more prescriptive when designing the search functionality.

Approaches to search embedded in clinical systems need to be tailored to the way different groups of clinicians work and think. While this guide identifies useful techniques to support searching, the ways in which these are applied may need to vary to match the working practices of a clinical specialty or the priorities of a particular health provider organization.

For effective data entry, the interface needs to be designed to meet the requirements for subsequent display and retrieval. Data entered and stored in the clinical record, needs to be displayed in ways that can be easily read and

understood. The data should also be retrievable to enable reporting analysis, decision support and appropriate communication to other systems and other users.

## 2.2. Using SNOMED CT Features to Support Optimized Searches

The logical model of SNOMED CT is able to support development of search techniques which can support effective and efficient Concept and Description retrieval.

It is however important to be aware that different techniques apply to different types of searches in order to avoid unconstrained searches which can make browsing SNOMED CT like navigating a jungle. This section presents some features of SNOMED CT which can help support effective and efficient searches.

### Benefits of Using SNOMED CT Hierarchies

The `|is a|` hierarchies in SNOMED CT support searches to be constrained by type, e.g. searches for Concepts which are contained in one of the top-level hierarchies such as `|clinical finding|`, `|body structure|` and `|procedure|` (see 4.4.2. [Constrain Searches by Supertype Ancestors](#)).

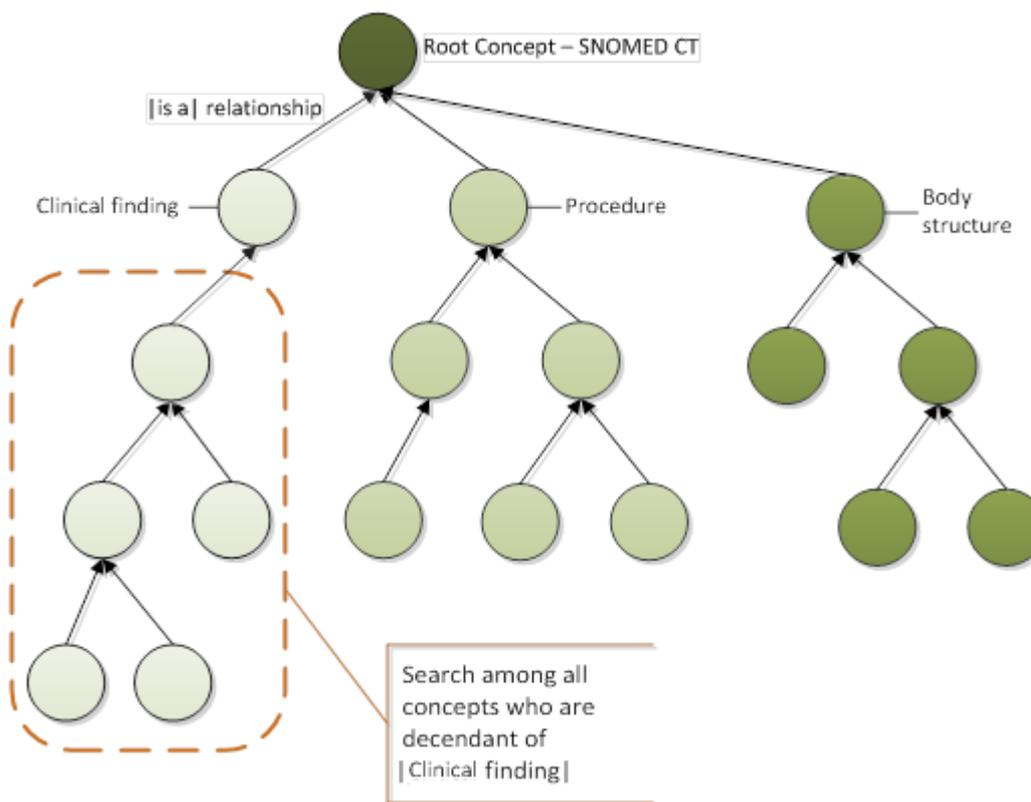
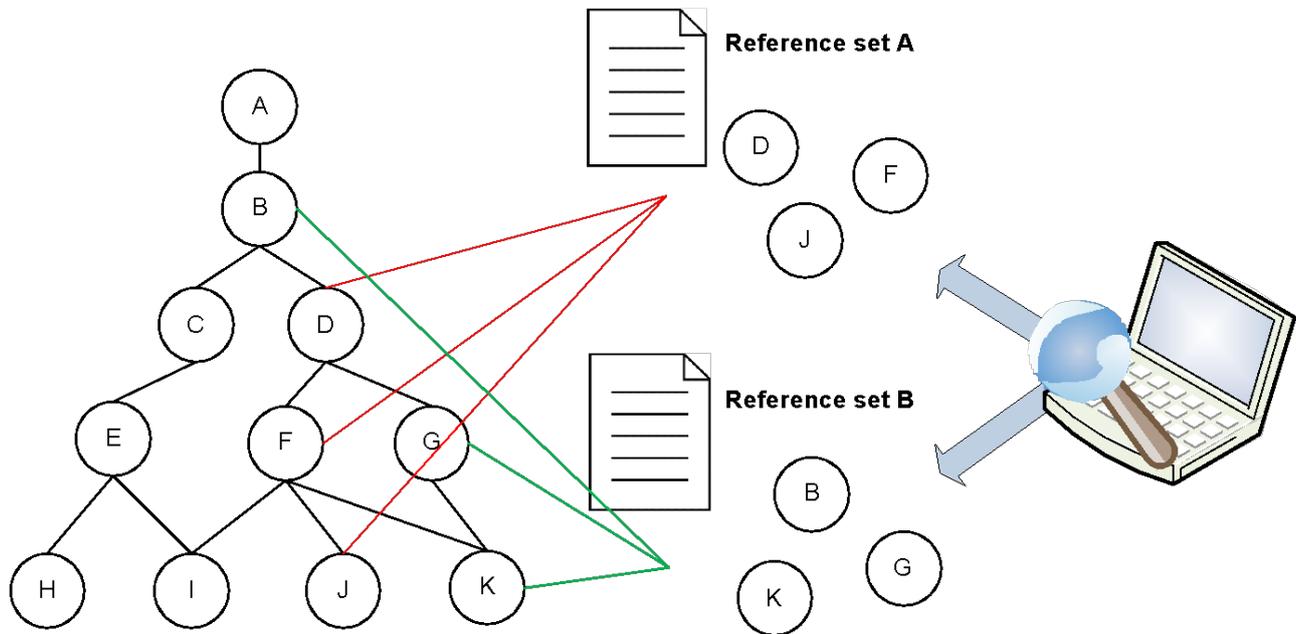


Figure 2.2-1: Search by subtype hierarchy

This type of search limits the number of search candidates significantly, as it excludes all Concepts of types other than those specified by the chosen subtype hierarchy. More granular Concepts than the top-level Concepts can be used as the basis for this type of search. For instance, searching for subtypes of `|Cardiovascular finding|` will further concentrate the type and number of candidate results.

## Benefits of Using Reference Sets

SNOMED CT supports the definition of Reference Sets. Reference Sets are important as they can be used to constrain search and data entry or support SNOMED CT navigation. Hence, searching within a Reference Set will limit the number of search results and ensure that the result set will be relevant for the desired context (see [4.4.3. Constrain Searches by Reference Sets](#)).



**Figure 2.2-2: Search using Reference Sets**

Moreover, Reference Sets can also be used to exclude Concepts as possible result candidates. For example, as SNOMED CT contains both animal and human Concepts, for human medicine it is likely that Concepts in the non-human Reference Set needs to be excluded from all searches.

## Language Reference Sets

It is a benefit that SNOMED CT supports development of Language Reference Sets, as this supports searches by Descriptions that are limited to a specific language or dialect (see [4.4.4. Constrain Searches by Language or Dialect](#)). A Language Reference Set identifies the Descriptions which are considered to be preferred designation for and acceptable as alternatives for SNOMED CT Concepts in the context of a given language and, optionally, dialect.

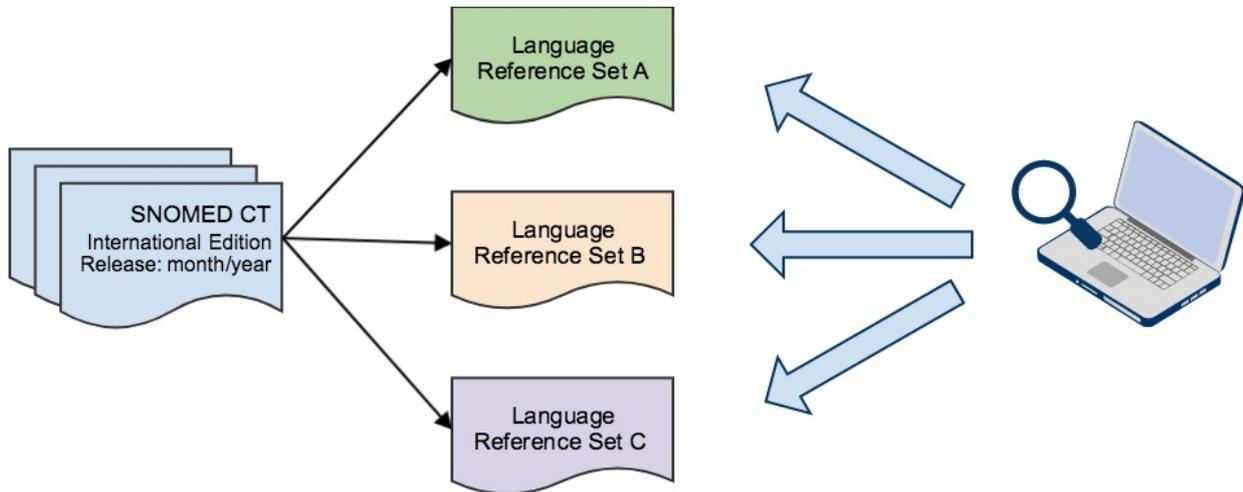


Figure 2.2-3: Search in different languages using language reference sets

## 3. Use Cases for Searches

Searching for a Concept can satisfy a wide range of use cases. This section describes common use cases that require effective search techniques. Each of these use cases may require a different combination of search techniques. The first sub-section describes SNOMED CT within the context of a clinical record application which is directly connected to data entry, and the second sub-section describes the use of SNOMED CT within a browser.

### 3.1. Use Cases Directly Connected to Data Entry

#### Writing an Entry Into a Clinical Record at or Near the Point of Care

At the point of care, search is done in order to find a Concept which describes a specific event that occurred during the course of patient care (e.g. documenting the findings related to a physical examination, diagnosis or surgical procedure into a form).

#### Creating Forms for Structured Data Entry

Data entry forms in clinical records are often designed for a specific documentation purpose. Designing the interface of a data entry forms and/or its content may entail associating Concepts with a template element or control. In highly structured forms, the design may involve specifying what Concepts should be available for selection. In less structured templates, the design may involve specifying the features useful for conducting an efficient and effective search. Thus, template designers may need to be able to search the whole terminology but also have the competencies to manually constrain the search by various filters and applying advanced functionalities, such as expression views and comparison between Concepts.

For more information on structured data entry, please refer to the "Structured data entry" sub-section.

### 3.2. Use Cases Where Search Browsers are Required

#### Creating Queries for Retrieving Data

Search can be required for defining Concepts and Descriptions that are used in subsequent data retrieval from patient records. The effectiveness of a search technique depends on the requirements for reporting and analysis (e.g. creating a query that searches for SNOMED CT coded data within specific subtype-hierarchies).

#### Binding SNOMED CT to Knowledge Artifacts

The integration of clinical decision support tools into clinical record systems can help to improve evidence-based practice. The binding of SNOMED CT to knowledge artifacts enables the linkage between the knowledge content in clinical decision support and the SNOMED CT coded data items in the clinical record. The binding typically entails searching for the SNOMED CT Concept. Effective techniques for searching for SNOMED CT Concepts for data entry can improve the assessment of the context of a Concept within SNOMED CT prior to binding and the overall binding efficiency.

#### Evaluating Terminology Content

*For example to review content coverage*

There can be various use cases for reviewing SNOMED CT content. For example, terminology content can be evaluated to decide whether the coverage of SNOMED CT is a sufficient to fit documentation purposes (e.g. a cardiovascular surgeon could assess whether SNOMED CT covers all cardiovascular terms). This assessment typically entails searching for SNOMED CT Concepts which requires effective search techniques.

## Mapping Between SNOMED CT and Another Code Scheme

In mapping, searching to identify SNOMED CT Concepts required to be mapped to or from another code scheme (e.g. ICD-10 or local codes) is required.

## Creating Reference Sets

*For example to represent subsets of terms and concepts*

Identifying subsets of Concepts or Descriptions for inclusion in a Reference Set typically requires searching SNOMED CT content.

## Learning About the Structure and Content of the Terminology

Realizing and understanding the comprehensiveness and complexity of SNOMED CT may be a steep learning curve for users. Searching and browsing through navigational results can aid the learning process. Thus, effective search techniques increase the chances of finding the required content and understanding where it fits into the overall structure of SNOMED CT.

## Editing SNOMED CT Descriptions

Terminology authorities such as National Release Centers may need to edit SNOMED CT Descriptions to adhere to national or local linguistic/editorial guidelines. This entails searching large subsets of Concepts and Descriptions. Thus, effective search techniques are required to find the correct Concept or Description within these large subsets.

## 4. Optimizing Searches

Today, a variety of general purpose text search tools and software libraries are in everyday use. Many of these general techniques can be applied to SNOMED CT though this does not automatically mean that they should be or that they are sufficient. Optimal indexing of SNOMED CT Descriptions may require a customized indexing solution. There are several widely available SNOMED CT browsers and related search tools, some of which use general purpose search techniques, and others which use some of the specific features discussed in this section. However, SNOMED CT searches often result in a very long list of terms, complete with additional information which must be assessed to fully understand the difference between terms in the search results. For example, searching for "diabetes" in most browsers returns a long list of lexical matches from multiple sub-type hierarchies. Choosing the appropriate term from this list can be a challenge if the terminological context is not apparent from the list.

Effective implementation of SNOMED CT depends on the speed and simplicity with which users can locate the terms and Concepts that they wish to use. A busy clinical user may become frustrated if the content they need cannot be quickly located when they search using familiar words or phrases. For this reason an efficient search strategy should address the many of the following key issues:

- Search should not be too sensitive to word order or exact phrasing:
  - Search should be insensitive to word - order variants:
    - For example, "head pain" for | pain in head |
  - Allow use of acronyms or abbreviations for frequently used terms:
    - For example, "MI" for "myocardial infarction" or "mitral incompetence".
- Excessive search results should not hinder selection of the required Concept:
  - When several synonyms of the same Concept match the search key, only one may need to be displayed.
- Speed result ordering
- Speed of search:
  - Search speed should be optimized
    - For example, use appropriate indexes (e.g. single key index).

The purpose of this section is to describe various strategies that might be used to implement the search requirements outlined above. A balance must be kept between the completeness and specificity of searches. Therefore, the next two sections consider techniques that extend searches to improve completeness and techniques that constrain searches to improve specificity.

## 4.1. Search by Text

### 4.1.1. Search by Text Containing Diacritic and Accented Characters

Diacritic and accented characters should be properly retrieved and displayed to ensure the completeness of a search. For instance, Sjögren must also be found by 'sjogren' and Ménière by 'meniere'). This example is specific for searches in English or in languages that do not use special characters. In countries, such as Swedish and Danish, that use special characters, it would be wrong to return "Sjögren" when searching from "Sjogren".

### 4.1.2. Search for Descriptions that Begin with the Search Text

This technique enables the user to find Concepts and Descriptions that begin with the text string entered in the search box. This search potentially reduces the number of keystrokes required to enter the search string. This technique is also used in scenarios where the user needs to find Concepts where they know the start of the description, but not the entire description. It is maybe, for many uses cases, a very useful technique to be configured as a pre-selected option that is user configurable.

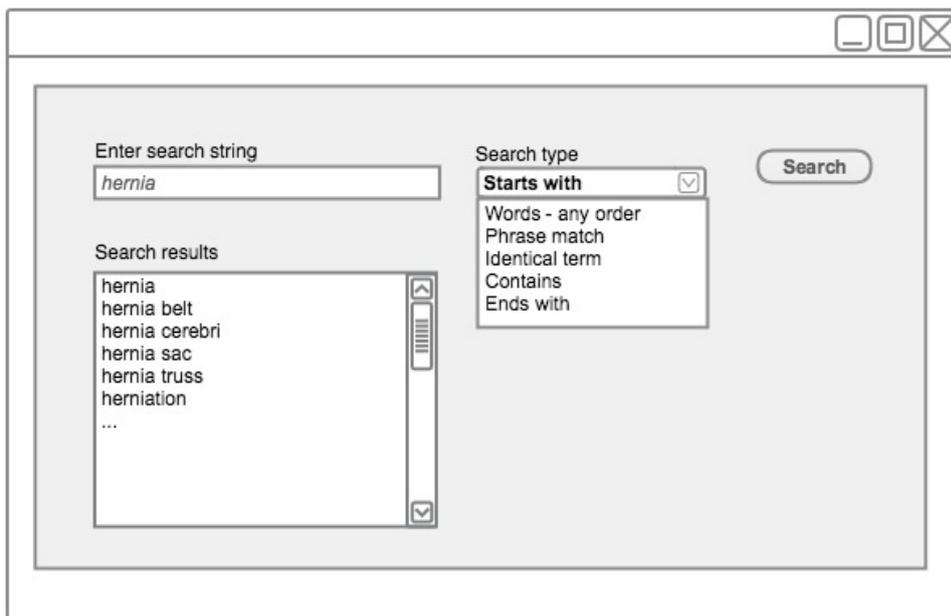


Figure 4.1.2-1: Search for descriptions that start with "hernia"

### 4.1.3. Search for Descriptions that Contain the Search Text

This technique may be used in scenarios where a user needs to find the required text anywhere in the term string including in the middle of words. However, this technique may take a lot of processing power, which could potentially impact search performance or the overall system performance. A large number of results may be returned which contain non-useful fragments (e.g. 'ete'). On the other hand, using this technique may be useful in languages that use contractions, such as German and Dutch, and there may be scenarios where the user may want to extensively search for the required text anywhere in the term string including in the middle of words such as medical pseudo-Latin & Greek terms (e.g. 'gastroduodenostomy'). It is therefore useful to have this option user configurable and not have it configured as a default option.

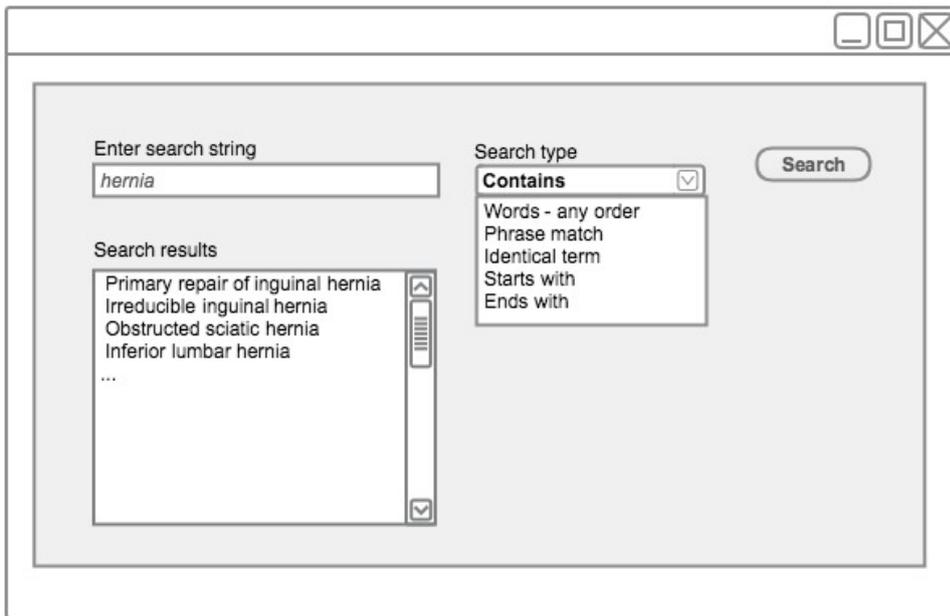


Figure 4.1.3-1: Searching for descriptions that contain "hernia"

#### 4.1.4. Search for Descriptions that End with the Search Text

This technique enables the user to find Concept and Descriptions that end with word ending(s) entered in the search box. This is not considered to be a very useful technique for many use cases, nevertheless, there may be scenarios where the user may want to search for the required text at the end of words. E.g. drugs ending in "statin" – finds Atorvastatin, Cerivastatin, Fluvastatin. It may be useful to have this option user configurable and configured as a default option.

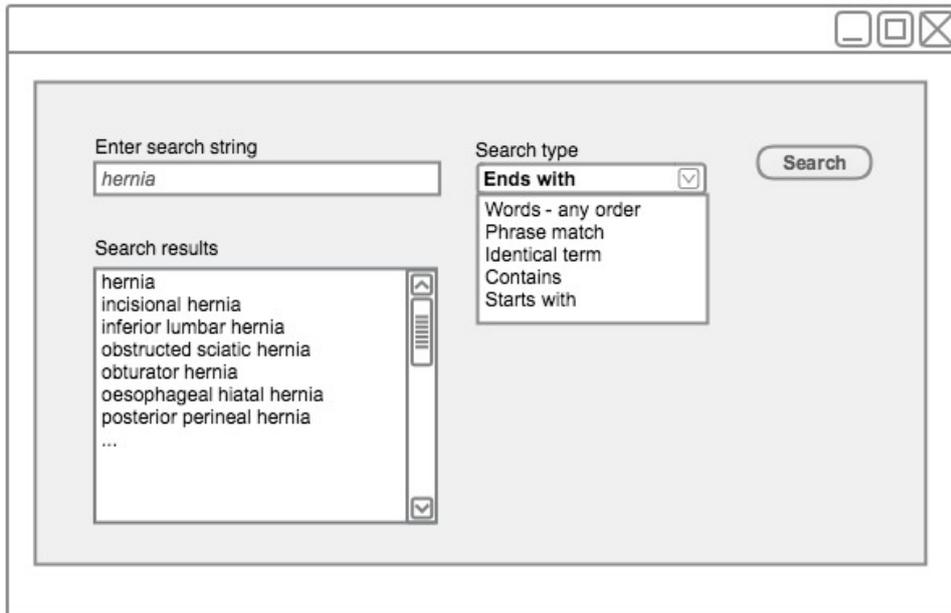


Figure 4.1.4-1: Searching for descriptions that end with "hernia"

### 4.1.5. Search for Words within in Any Order

This technique enables the user to find descriptions that contain the search text string(s), whether these are at the beginning, at the end or in the middle of a description. This search type would be useful for users who do not know how words are ordered in the descriptions. It may be useful to have this option user configurable and not have it configured as a default option.

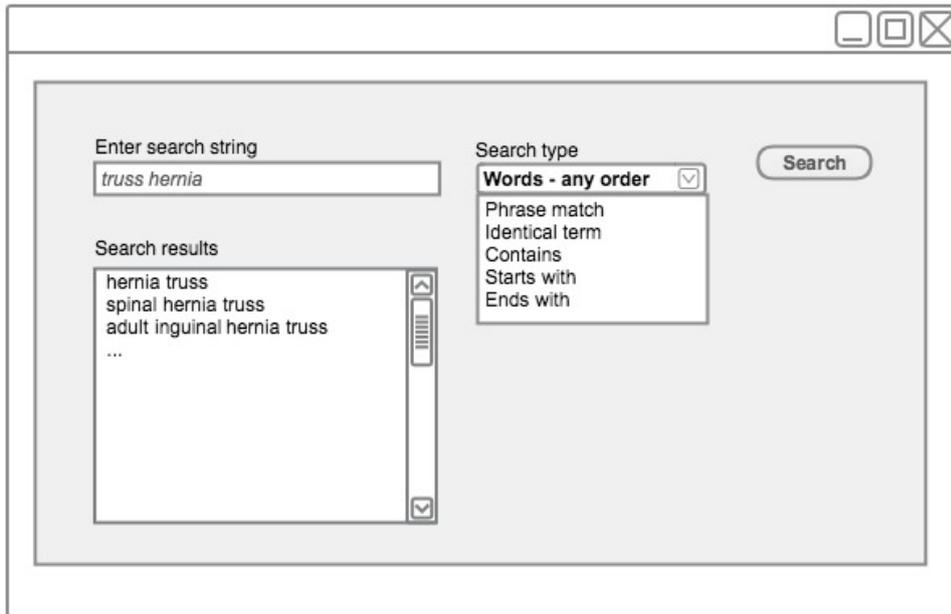


Figure 4.1.5-1: Searching for words in any order

#### 4.1.6. Search for Identical Terms

This technique enables exact matches to be found in the search result instantly regardless of the interposing words; otherwise, search on a single word may produce many matches. This technique should be used in scenarios where the wording of the desired description is known to the user who wants to search for and instantly select from the search candidates without having to scroll through a long list of search results. It may be useful to have this option user configurable and not have it configured as a default option as it may overly constrain the search leading to missing candidates.

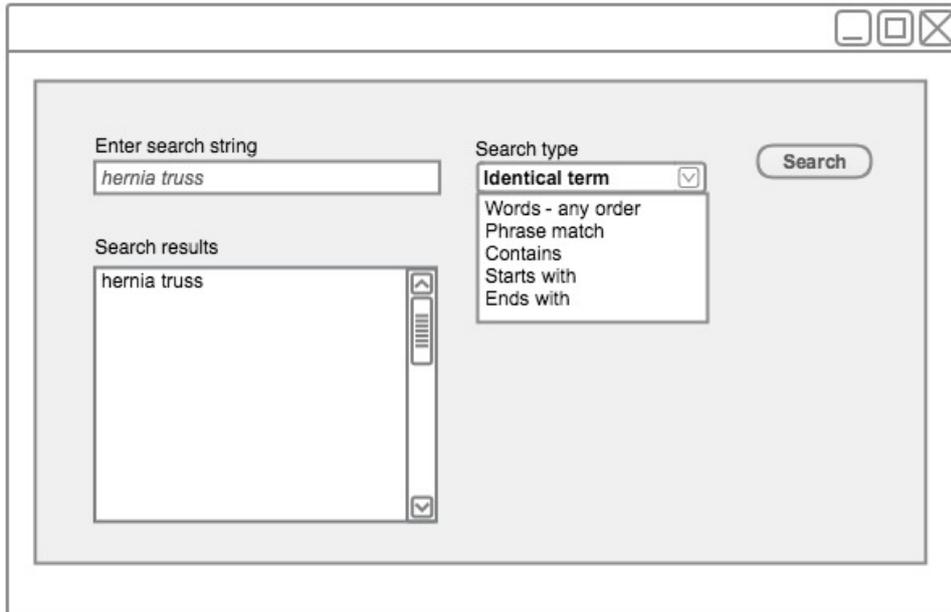


Figure 4.1.6-1: Searching for identical terms

### 4.1.7. Search for Words in a Specific Order or a Matching Phrase

This technique enables the user to find Descriptions that matches all words in the search text. This means that the order of terms in the Concept or Description should be in the exact same order as the terms fully or partially represented in the search text. It may be useful to have this option user configurable and not have it configured as a default option.

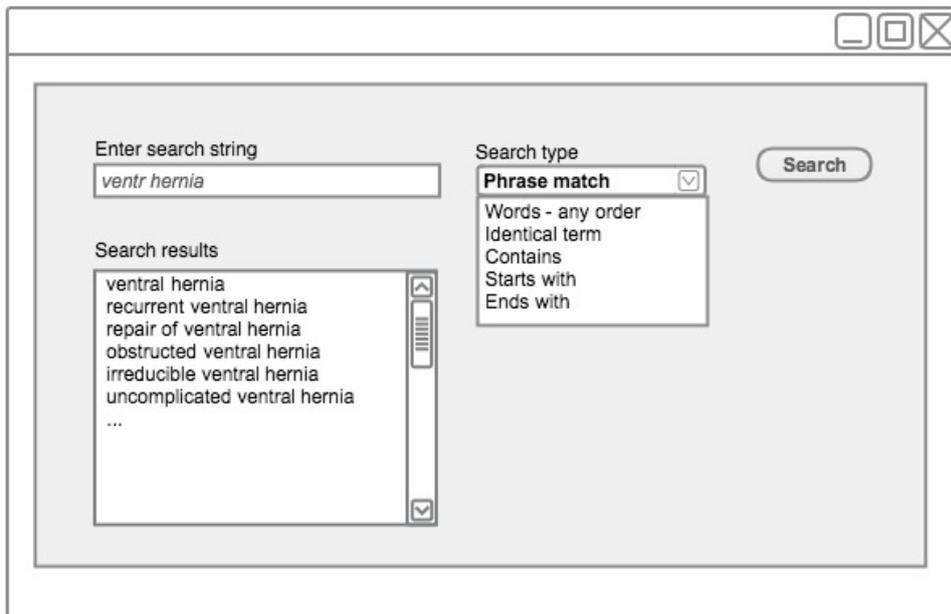


Figure 4.1.7-1: Searching for words in order or a matching phrase

## 4.2. Search by Identifiers

Searching by SNOMED CT identifier either as a Concept Identifier or a Description Identifier is useful in scenarios where the user knows the SNOMED CT Concept or Description they require.

It is useful in use cases for browsers, but is not an appropriate way to search in clinical applications. As a general rule the default behavior of clinical application should not display identifiers and users should never be expected to remember, type or cut-and-paste identifiers for data entry.

## 4.3. Extended Searches

Extended searches enable the search to return more candidate matches. The intention of extending searches is to maximize the likelihood that the required Concept or Description will be included in the candidate matches. This sub-section describes some recommended techniques for extending searches.

### 4.3.1. Extend Search by Word Equivalents

In healthcare, there are many words with equivalent meanings. Synonyms provide alternative phrases referring to the Concept. However, Synonyms are not created automatically for every possible combination of words with an equivalent meaning. The success of simple searches using one or more keywords depends on the text of the available Descriptions. Therefore searches will fail or will be incomplete where a different but equivalent word is used in the search.

SNOMED CT cannot and does not attempt to enumerate all possible *synonyms* for all Concepts. Neither to enumerate all possible semantically equivalent typographic, part-of-speech or ordering variations of the synonyms it does include. To do so, especially automatically, would both require a considerable lexical resource and hugely increase the number of Descriptions and the size of any associated index tables. An alternative technical architecture involves a selective, logical expansion of search expressions and the relevant fragment of Description-space at run-time, using substantially the same lexical resources

For example: | Kidney stone | and | Renal calculus | are synonymous Descriptions in SNOMED CT. However, a search of SNOMED CT for the target phrase "kidney stone fragmentation" yields the result | Percutaneous nephrostomy with fragmentation of kidney stone | while a search for "Renal stone fragmentation" yields no results. One way of addressing this problem is to maintain a table of *word equivalents*, and a table of this type is a prerequisite for exhaustive synonym generation. The Word Equivalent Table is included in the Developer Toolkit of SNOMED CT. [see [SNOMED CT Developer Toolkit Guide](#)]. Individual implementers will wish to add additional *word equivalents* to meet the requirements of their particular medical specialty or user needs. The resulting table then acts as an additional resource to assist searching and parsing of phrases. It need not be a comprehensive dictionary of words.

#### Example:

"Tap" and "aspiration" are equivalent in the context of *terms* such as "pleural tap", "pleural aspiration", but not in the context of a "patella tap", a physical "tap" on a bag or catheter, or the clinical disorder "neonatal aspiration syndrome".

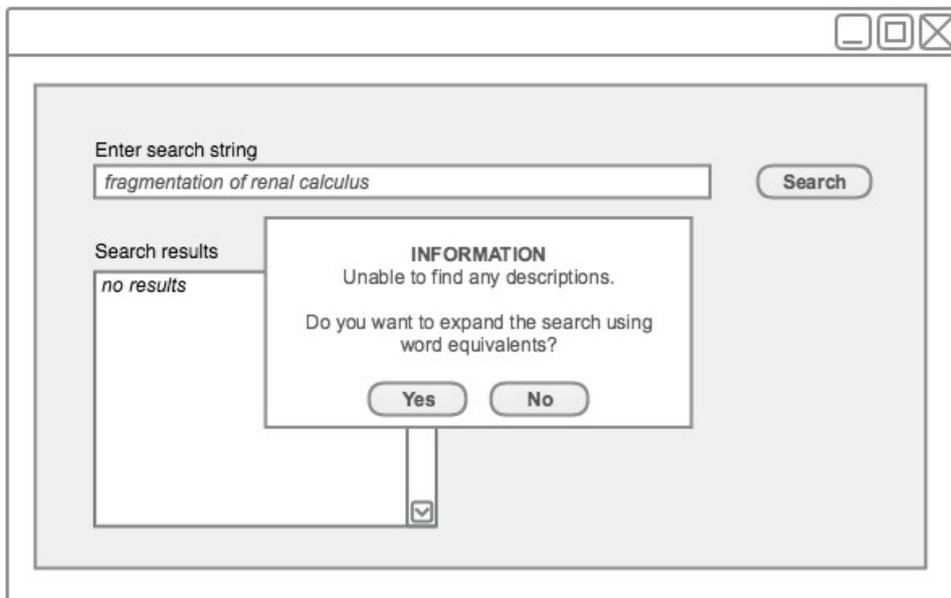
When searching using incomplete words and/or wildcards, use of *word equivalents* may impede effective searches by increasing the number of spurious potential matches. This either extends the processing required to filter the real matches from the potential matches or increases the length of the list of choices presented to the user.

#### Example: Using word equivalents table to extend a failed search

A system user enters the search string "Fragmentation of renal calculus;" the search returns no results. The search application that the user has been provided with has the option to extend the search by using the word equivalents table (see [Table 4.3.1-1](#))

**Table 4.3.1-1: Word Equivalents Table Example**

<b>WordBlockNumber</b>	<b>WordText</b>	<b>WordType</b>
1021	KIDNEY	2 (word equivalent)
1021	RENAL	2 (word equivalent)
4430	CALCULUS	2 (word equivalent)
4430	CALCULI	1 (word form variant)
4430	STONE	2 (word equivalent)
9870	RENAL STONE	4 (equivalent phrase)
9870	KIDNEY STONE	4 (equivalent phrase)
9870	KIDNEY CALCULUS	4 (equivalent phrase)
9870	RENAL CALCULUS	4 (equivalent phrase)
9870	NEPHROLITH	2 (word equivalent)

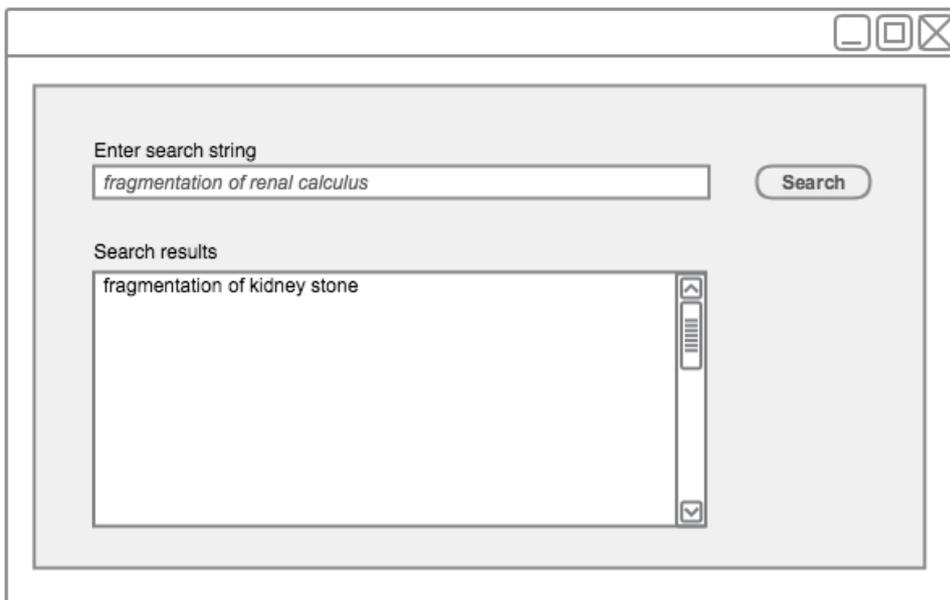


**Figure 4.3.1-1: Extending searches with word equivalents - step 1**

The user selects this option (or it is invoked automatically because the initial search without word equivalents found nothing) and searches again using the same search *string*. The table is used to make substitutions in the search *string* to produce all possible unique search variants:

- "Fragmentation of renal calculus"
- "Fragmentation of renal stone"
- "Fragmentation of kidney stone"
- "Fragmentation of kidney calculus"
- "Fragmentation of Nephrolith"
- "Fragmentation of renal calculus"
- "Fragmentation of renal calculi"
- "Fragmentation of kidney calculi"

These 8 search variants are used as the target phrases for the searches. The search results of these variants are combined, duplicate Concepts are eliminated and the list of search results is returned.



**Figure 4.3.1-2: Extending searches with word equivalents - step 2**

### 4.3.2. Extend Search by Postcoordinated Searching

A SNOMED CT search can be expanded to support appropriate or commonly used qualifiers. This technique is useful in situations where searches fail to find a precoordinated Concept. This is particularly useful in clinical applications which support the storage of postcoordinated expressions. This technique is likely to increase the ease of entry of postcoordinated expressions in the clinical record as well as the overall usability of search and data entry.

As a basic implementation, a limited facility for recognizing commonly qualifying words may be used. For example, Concepts such as | left |, | right |, | routine |, and | emergency | are applicable as qualifiers for some Concepts when not already included in a precoordinated Concept.

Example:

The phrase "emergency closed reduction fracture left femur" might list "closed reduction of fracture of femur, emergency, left" which refers to the postcoordinated expression 112777008 |closed reduction of fracture of femur| : 260870009 |priority| = 25876001 |emergency| , 272741003 |laterality| = 7771000 |left| .

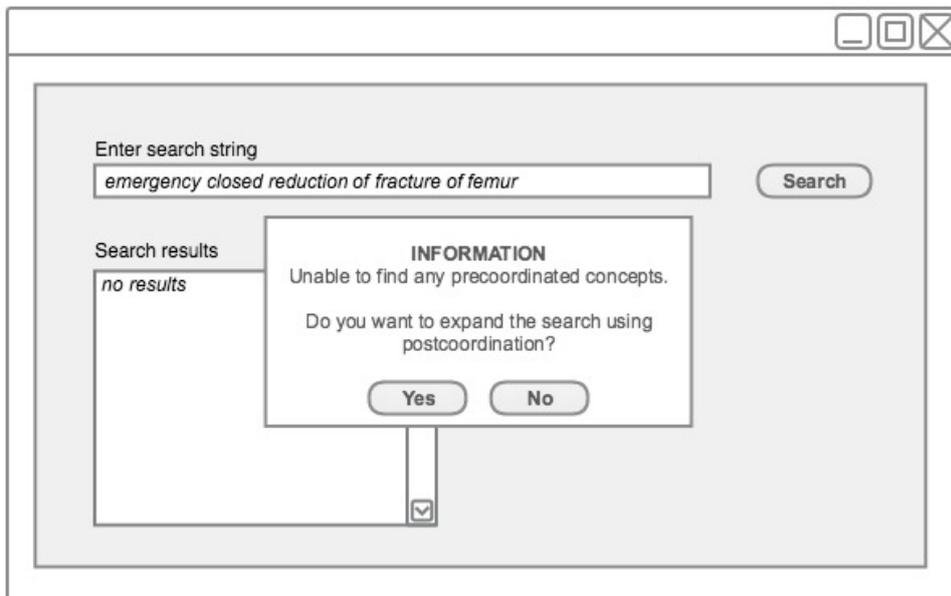


Figure 4.3.2-1: Extending searches by post-coordination - step 1

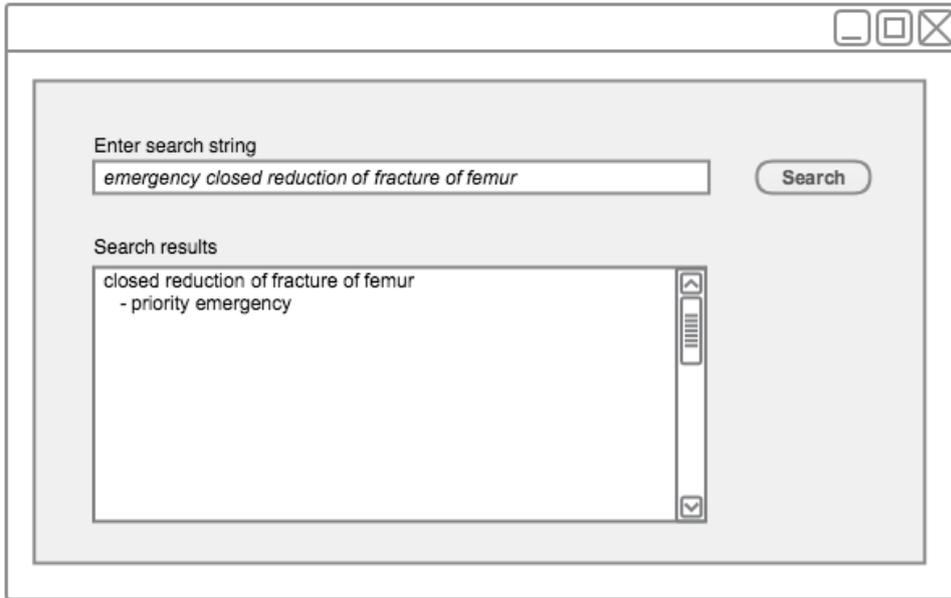


Figure 4.3.2-2: Extending searches by post-coordination - step 2

## 4.4. Constrained Searches

Constraining searches enables the search browser to return fewer candidate matches. This is opposite to extending searches (see 4.3. [Extended Searches](#)). The intention of constraining a search is to avoid getting a long list of search results to scroll through. This shortens the time the user needs to find the required Concept, thereby increasing the usability of the application.

### Constrain searches by excluding "stop words"

Frequently used words with limited semantic specificity can be excluded from search indexes to improve the speed and specificity of searches. Examples of English words typically included in "stop word" lists include: a, an, and, as, at, be, by ... of ... the, etc.

General purpose search software often includes a default list of "stop words" but these may not be entirely appropriate for SNOMED CT searches. For example, the MySQL default "Full-Text Stopwords" includes "no", "non", "not", "without" and many other words which may be highly significant when searching for clinical terms.

The SNOMED CT International Release includes a suggested "stop list" (referred to as the *ExcludedWords* Table). This is only available in English and is the list of exclusions used when generating keyword lists distributed with the same release.



**Figure 4.4-1: Constraining words in any order searches by excluding stop-words - step 1**

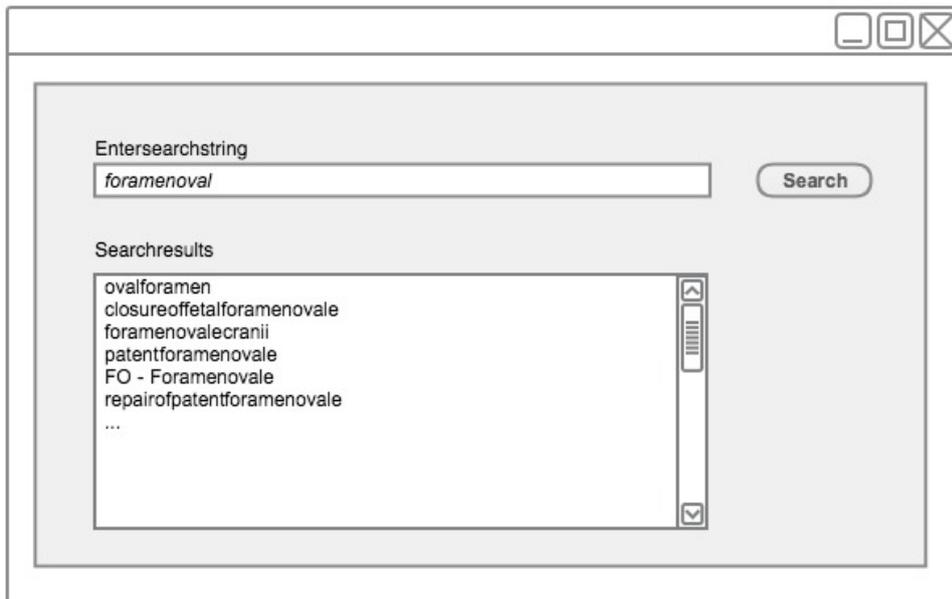


Figure 4.4-2: Constraining words in any order searches by excluding stop-words - step 2

### 4.4.1. Constrain Searches by Status

Searches should usually be filtered so that only Active Descriptions associated with Active Concepts are returned.

There are a few use cases where a user may legitimately wish to search *Inactive Concepts* and *Descriptions*. Possible cases include creating or editing queries that locate previously entered data recorded using *Concepts* and *Descriptions* that are no longer recommended for *active* use. Therefore, searches intended to be used for these cases should allow the default active status filter to be disabled.

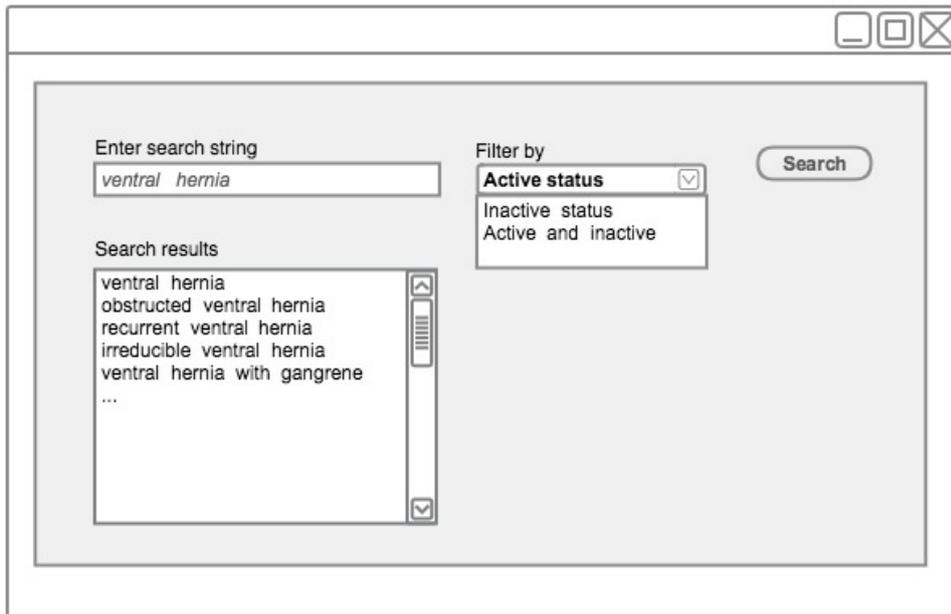


Figure 4.4.1-1: Constraining searches by filtering by active status

#### Important Note

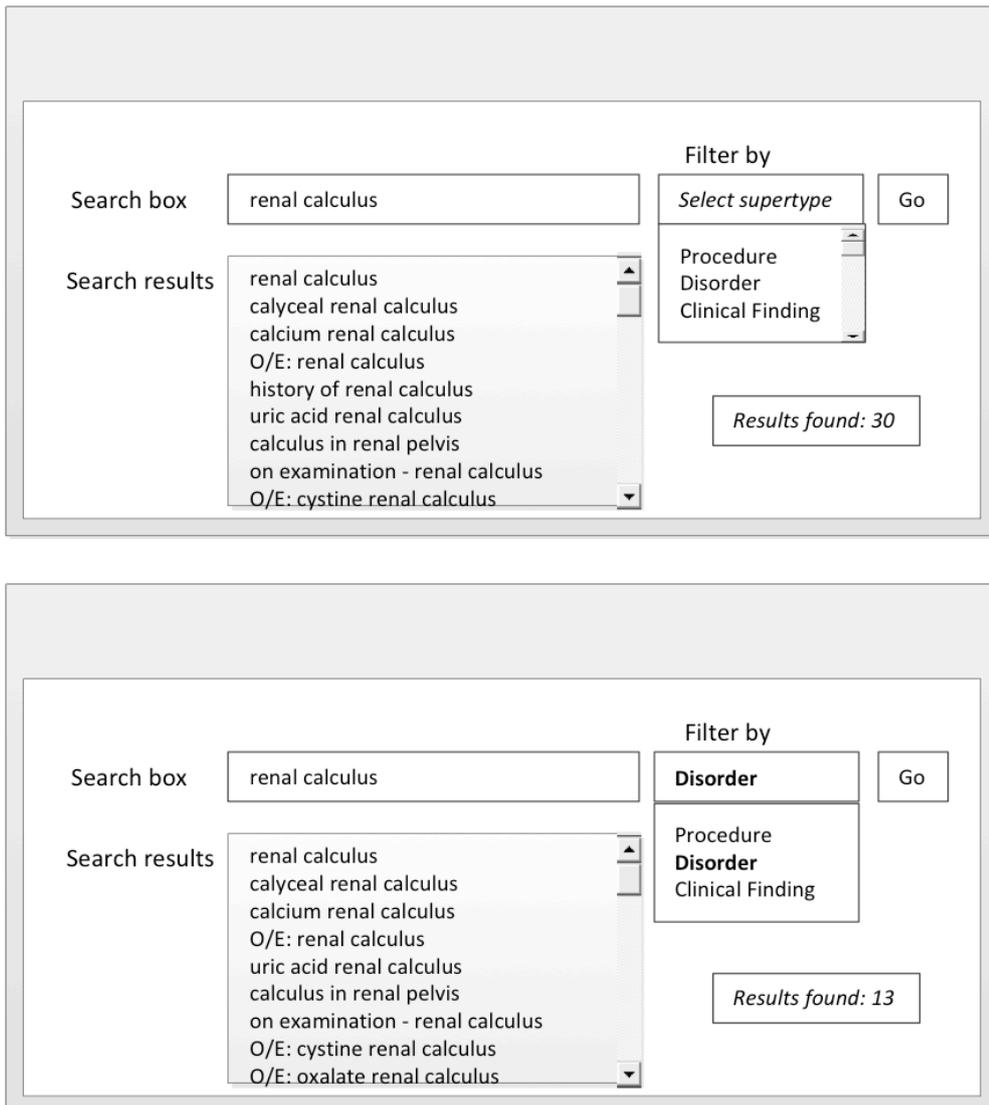
For use cases involving data entry or design of data entry template or Reference Set, this filter should always be applied and set as active. This should not even be a filter option.

## 4.4.2. Constrain Searches by Supertype Ancestors

Searches may usefully be limited to *Concepts* that have a specified supertype ancestor, which is appropriate for the context of a particular field, template or protocol.

### Example

When attempting to record the diagnosis "renal calculus," it is not helpful for a search to include the procedures that may be carried out to treat a renal calculus.



The figure consists of two screenshots of a search interface, illustrating how filtering by supertype ancestors affects search results.

**Top Screenshot:** The search box contains "renal calculus". The "Filter by" dropdown is set to "Select supertype". The search results list includes 13 items, including "renal calculus", "calyceal renal calculus", "calcium renal calculus", "O/E: renal calculus", "history of renal calculus", "uric acid renal calculus", "calculus in renal pelvis on examination - renal calculus", and "O/E: cystine renal calculus". A "Results found: 30" box is displayed.

**Bottom Screenshot:** The search box contains "renal calculus". The "Filter by" dropdown is set to "Disorder". The search results list includes 13 items, including "renal calculus", "calyceal renal calculus", "calcium renal calculus", "O/E: renal calculus", "uric acid renal calculus", "calculus in renal pelvis on examination - renal calculus", "O/E: cystine renal calculus", and "O/E: oxalate renal calculus". A "Results found: 13" box is displayed.

Figure 4.4.2-1: Constraining the search by filtering by supertype ancestors

### 4.4.3. Constrain Searches by Reference Sets

Searches for Descriptions or Concepts may need to be constrained by Reference Sets. Applications should allow searches to be filtered, ordered or otherwise prioritized in accord with one or more active Reference Sets. Specifically, the search mechanism should support the following functions with respect to the following types of Reference Sets:

- A Simple Reference Set may be used to filter, sort or highlight the results of text search or hierarchical navigation. This may simplify or encourage selection of Concepts or Descriptions used in a particular country, organization or specialty.
- A Simple Reference Set or an Ordered Reference Set may be used to specify or order the valid Concepts for entry in a particular field.

Reference Set may also be used to exclude subsets of Concepts and Descriptions that are not relevant or appropriate in browsers and in-built search functionalities. For example, it may be beneficial for the search not to return any members of the non-human Reference Set, unless the search is explicitly customized to include them.

### 4.4.4. Constrain Searches by Language or Dialect

Constraining searches by language or dialect is a type of constraining by Reference Sets. For every translated version of SNOMED CT, this technique is mandatory. Filtering of search and navigation results to include only those Descriptions that are referenced by the Language Reference Sets may be applied to limit a search to those Descriptions applicable in a particular language or dialect. A single common encoding language should be mandatory in search use cases directly connected to data entry, to avoid confusion and ambiguity. There are a few use cases where a user may legitimately wish to search for Descriptions included in extensions of other countries. For example, when mapping SNOMED CT to other code sets, en-US (American) Descriptions may be required if extensions are not available in the UK language Reference Set.

## 4.5. Improve Search Speed

The intention of improving search speed is to improve the usability of data entry. This sub-section describes techniques that can help speed up searches.

### 4.5.1. Enable Real Time Searching

Real time searching offers suggestions to finish the words being typed by the user or displays the results while the user is typing. It can improve the user experience by reducing the number of keystrokes a user has to make and to help reassure them that the system 'understands' their intentions.

Conventional text searches require the user to decide how many words to enter and then explicitly request a search. When a search fails to find any matches or returns a very long list of matches, the user is obliged to repeat the process. The need to undertake this type of user interaction for every coded entry is likely to create a significant disincentive to effective data entry.

While auto-completion may generally be a useful technique, it may not be useful for data entry at or near the point of care as it may be distracting to the user and could potentially slow down the clinical documenting. It also requires the user to pay close attention to ensure that the auto-completed phrase is the one that the user intended to capture.

### 4.5.2. Show an Indication of Estimated Number of Matches Before Starting a Search

One possible solution to show the number of search results is an interface that performs real-time checking of the number of search results while the user is typing. The interface may indicate this to the user, allowing them to decide when to stop typing and commence the search. A further enhancement is to automatically return the list of matches whenever the user stops typing, or when the number of matches is reduced to an acceptable level.

### 4.5.3. Allow Slow Searches to Be Paused or Cancelled

When a search is slow to find any matches, it may be useful to give the user the option to pause or cancel the search. This technique would benefit applications or scenarios which either require:

- Less frequent data entry;
- Searching a large SNOMED CT subset;
- A lot of processing power, which could potentially impact search performance or the overall system performance, or
- Access to a terminology server via an Application Programming Interface (API).

### 4.5.4. Enable Background Encoding

Techniques that support real-time searches [see 4.4.1 Enable real time searching] and qualifier resolution [see 4.2.2 Extend search by post-coordinated searching] may also be extended to enable background encoding of complete sentences as they are entered. This method can be applied to text entered by typing or by voice recognition.

As text is entered, the search mechanism attempts to narrow the selection. If there are multiple possible matches, then these are presented to the user for selection. If there is a single match, the encoding is accepted by default; otherwise the user can override the match by selecting an alternative Concept or Description. A single good match is found at the end of the process, thereby encoding the text in the sentence.

There are many possible variants on this technique. For example, as the possible matches are narrowed down, the system could offer an auto-completion option similar to that used in web browsers and word-processors.

**! Caution**

Anyone implementing this approach should take care to undertake appropriate quality assurance of the results. Mention of this approach to data entry does not imply that it is considered safe for a given use-case. Formal professional assessment of the risks and benefits of any type of automated encoding is essential.

#### 4.5.5. Enable Automatic and Semi-Automatic Encoding

Techniques similar to those used for background encoding can be applied to previously entered text or to text entered by voice recognition or optical character recognition. Where such methods are used there is likely to be a need for manual intervention to resolve uncertain encoding. The requirement for manual intervention will depend on the sophistication of the matching techniques and the extent to which accuracy is safety-critical. If encoded data is to be used by clinical decision support protocols, which may influence the treatment of a patient, extreme care is needed when using automatic encoding and tools that allow manual review are essential. A less rigorous approach may be acceptable where the purpose of encoding is for aggregation and analysis of large volumes of population data.

**! Caution**

Anyone implementing this approach should take care to undertake appropriate quality assurance of the results. Mention of this approach to data entry does not imply that it is considered safe for a given use-case. Formal professional assessment of the risks and benefits of any type of automated encoding is essential.

#### 4.5.6. Optimize Indexing

Indexing SNOMED CT content facilitates fast and accurate text searches. Search functionality queries the indexed content (by keyword) to return a list of search candidates matching the query. General purpose index features may be sufficient for enabling rapid SNOMED CT searches. However, developers and users should be aware that search tools may need to be configured for SNOMED CT searches due to different stop word indexes and to ensure effective searches of the types described in this guide.

See [Appendix A - Using Search Index Files in the SNOMED CT Toolkit](#) for further information.

## 5. Optimize Display of Search Results

This section describes how search results can be ordered rationally. When developing browsers and search functionality within clinical applications, careful consideration must be taken when specifying the prioritization of search results as adopting many of the techniques below may compromise the speed of search. The benefits of presenting the results in the best manner should therefore be evaluated against the speed of the search.

### 5.1. Order Search Results Rationally

This section concerns the rational ordering of search results. When developing browsers and search embedded functionality within clinical applications, careful consideration must be taken in specifying the prioritization of search results, as adopting many of the recommended techniques below may compromise the speed of search.

#### 5.1.1. Order Shortest Matching Terms First

Ordering the shortest matching term first is a common display requirement for search browsers and search functionalities in clinical applications. This applies text search techniques previously in the Guide (e.g. searching for descriptions that contain the search text). [see 4.1.1 Search by text] Many users can expect this technique to exist in all use cases concerning searches that return large result sets. It is intuitive to display the shortest term that is also the closest lexical match first. If the shortest matching term is not in the first set of visible matches, a user is likely to assume that there are no relevant candidate matches or use the wrong concept or description.

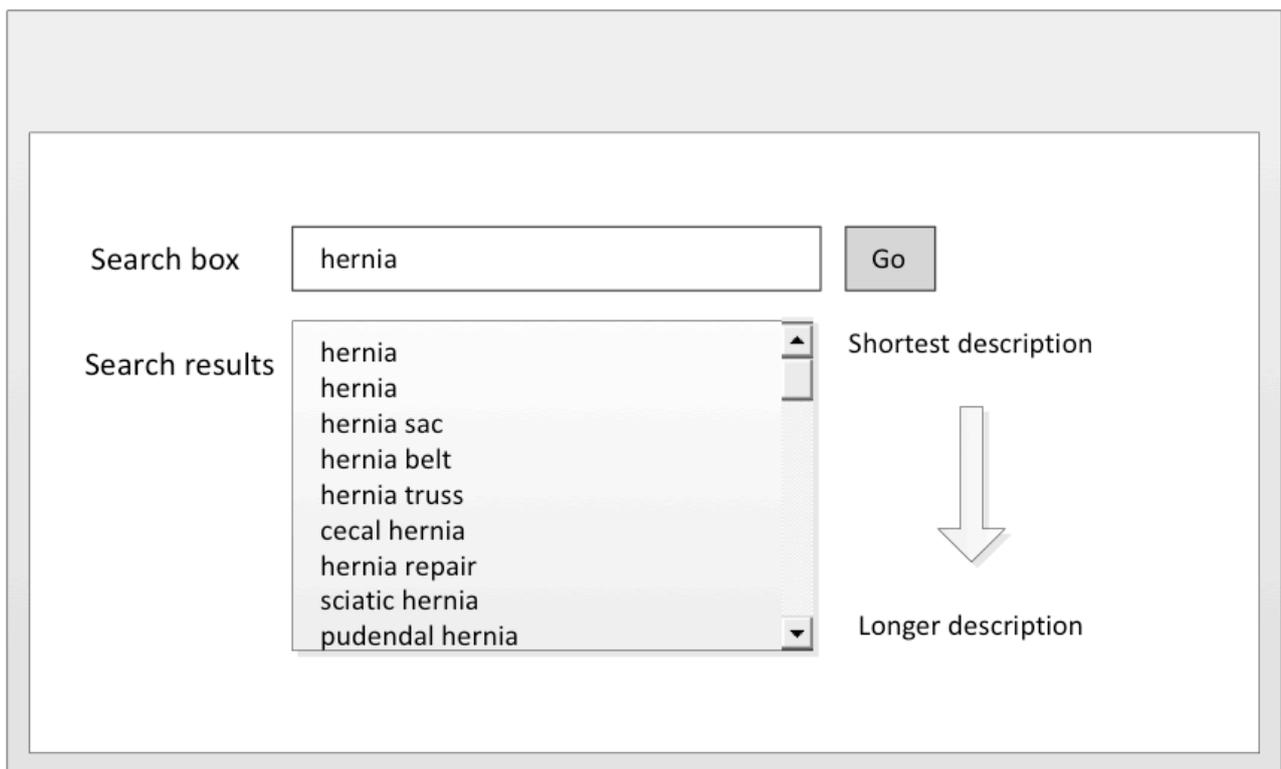


Figure 5.1.1-1: Ordering shortest matching terms first

A common mistake is to implement a search functionality that is configured to sort search results alphabetically in a clinical setting. The result list, when Concepts like hernia is searched, starts as shown below and the term "hernia" itself would be more than 130 items down a list of over 700 matches.

**Table 5.1.1-1: Position of term 'hernia' in an alphabetic and shortest terms ordered search 'hernia'**

Position	Alphabetic order of results	Shortest match first
1	abdominal hernia	<b>hernia</b>
2	abdominal wall hernia procedure	hernia sac
3	airway device cuff herniation	hernia belt
4	anesthesia for hernia repair in lower abdomen	O/E - hernia
...	anesthesia for hernia repair in upper abdomen	cecal hernia
6	anesthesia for lumbar or ventral incisional hernia of upper abdomen	labial hernia
7	anesthesia for transabdominal repair of diaphragmatic hernia	hernia repair
8	anesthesia for ventral or incisional hernia repair, lower abdomen	littré hernia
9	anterior perineal hernia	Cooper hernia
....		
<b>131</b>	<b>hernia</b>	

**! Caution**

Alphabetic searches can compromise patient safety as the best matching term is not easily found. However, even searches for the shortest term need to be constrained to concept types relevant to the clinical setting to ensure only relevant matches are shown.

### 5.1.2. Order Preferred Term Matches Before Synonyms

It is recommended to order preferred term matches before synonyms, particularly in clinical settings, as the *Preferred Term* is a common word or phrase used by clinicians to name that *Concept*. This technique will enhance usability and significantly increase the speed of data entry. It may be helpful to the user whether the candidate match is a Preferred Term or synonym.

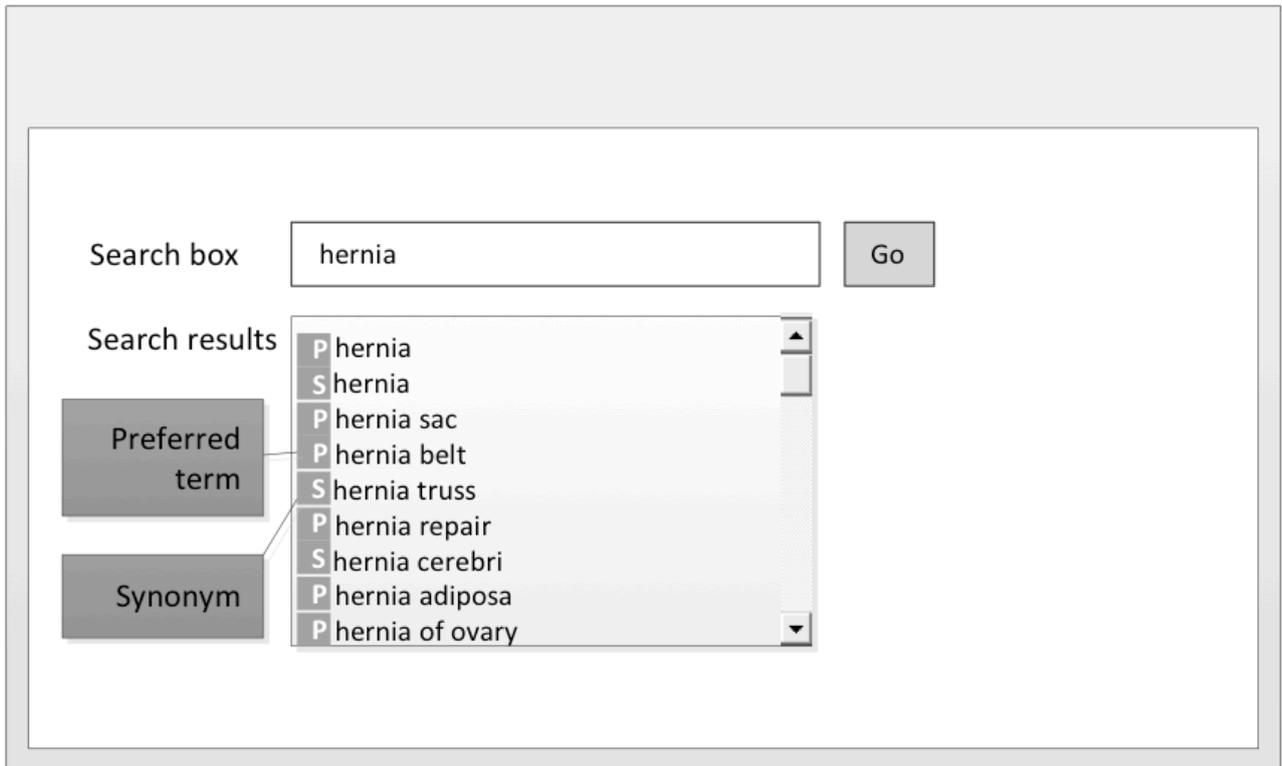


Figure 5.1.2-1: Ordering preferred term matches before synonyms

### 5.1.3. Order User Preferred Language Matches First in Multilingual Environments

For a country or region that has more than one official language (e.g. Belgium), it would be useful to have a user configurable option that enables results to display matches in a preferred language first in applications that support more than one language. This can be done using combinations of Language Reference Sets.

### 5.1.4. Order According to Priority in Any Active Reference Sets

Search results may need to be configured in a specified order using one or more active Ordered Reference Sets. The way in which access to search results is prioritized depends on the nature of the application and its operating environment. Examples of prioritization include:

- Showing Descriptions associated with high priority Concepts before those with lower priority.
- Showing Concepts with high priority before their less highly prioritized siblings in hierarchical display results. [see 5.5.3 Using the navigation hierarchy]
- Initially listing Concepts and associated Descriptions with priority above a specified threshold and requiring additional step to access those assigned lower priority.

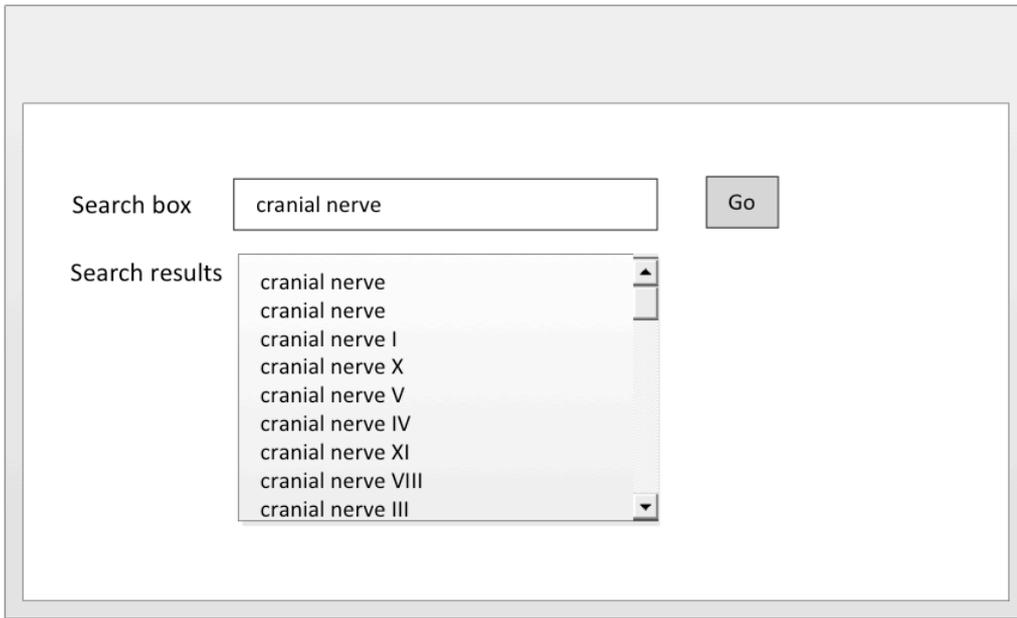


Figure 5.1.4-1: Unordered search results for cranial nerve

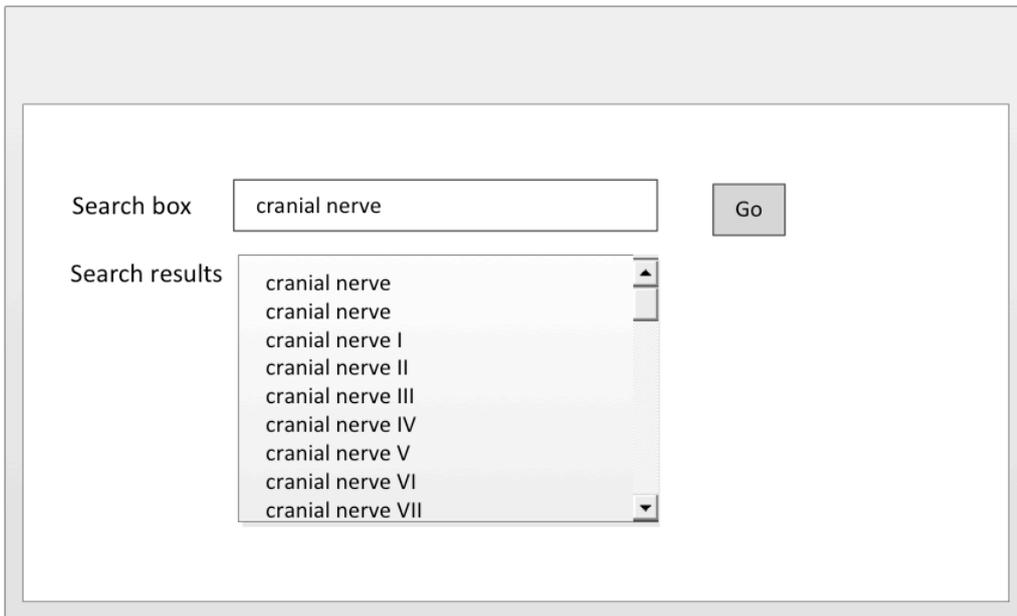
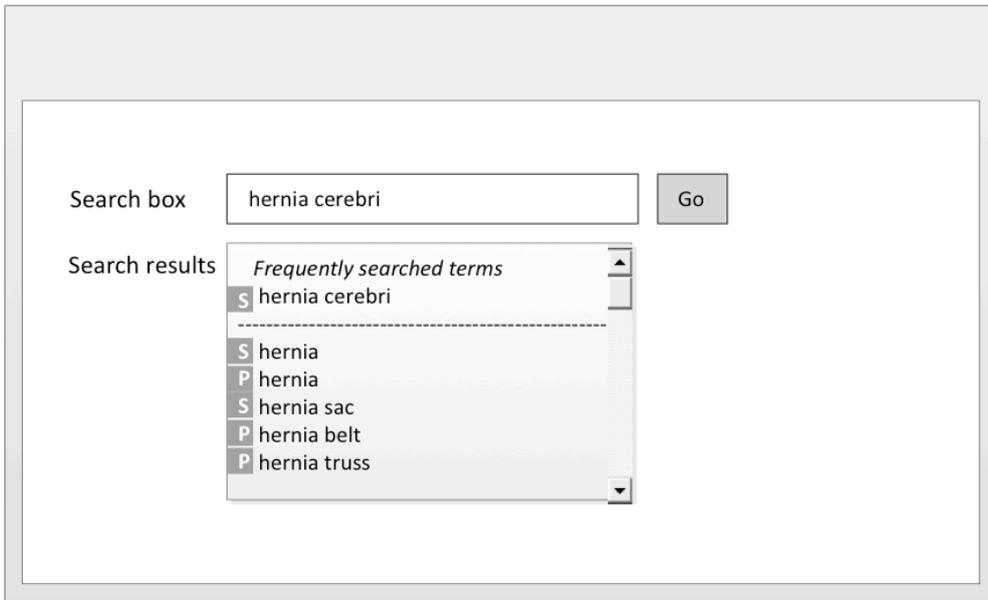


Figure 5.1.4-2: Ordering search results for cranial nerve using an ordered component reference set

### 5.1.5. Display the the Most Frequently Used Descriptions Listed First

This option will require the application to track the frequency of term selection so that search results can show the most frequently selected terms near the top of the results or displayed separately in the results. If this technique is chosen, user guidance may be necessary to minimize the risk of overusing the frequently used Descriptions as they would be easier to find. An alternative way of browsing through frequently used terms is to give the user the option to store the frequently selected terms as personal favorites and recall with user-defined abbreviated access [term](#) following a search. [see Use Mnemonics and personal favorites for data entry]



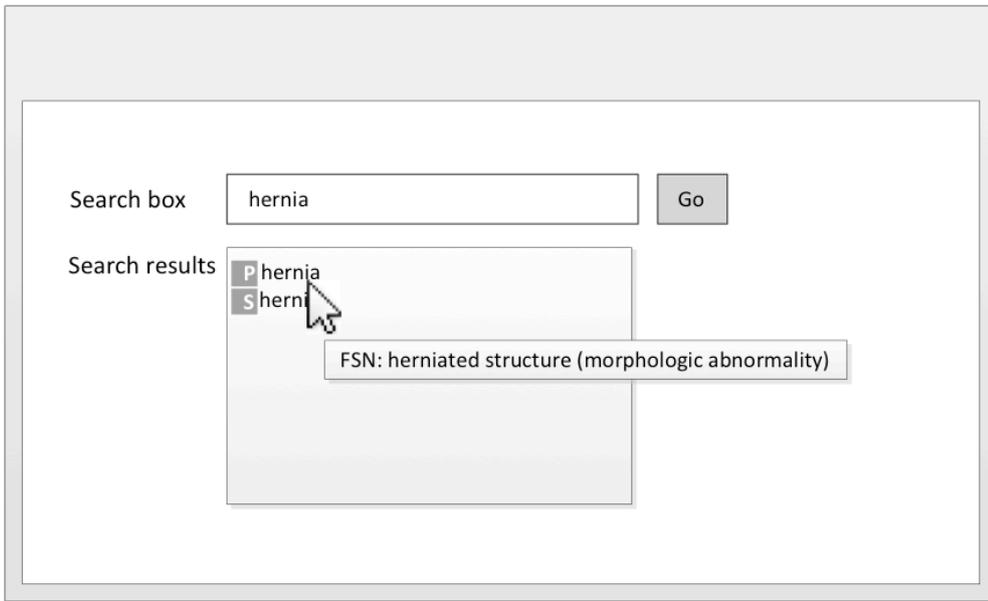
**Figure 5.1.5-1: Displaying most frequently used Description after this Description was entered in previous search**

### 5.1.6. Alphabetical Ordering

As noted in [5.1.1](#), alphabetical ordering of search results should **NOT** be used for general purpose searches of SNOMED CT. Instead the shortest matches, best matches or most commonly used matches should be shown first. However, there may be some specific situations in which it may be useful to alphabetically order a short list of matches. For example short pick-lists may be alphabetically ordered in a data entry form. [see Examples of applied mechanisms of structured data entry]

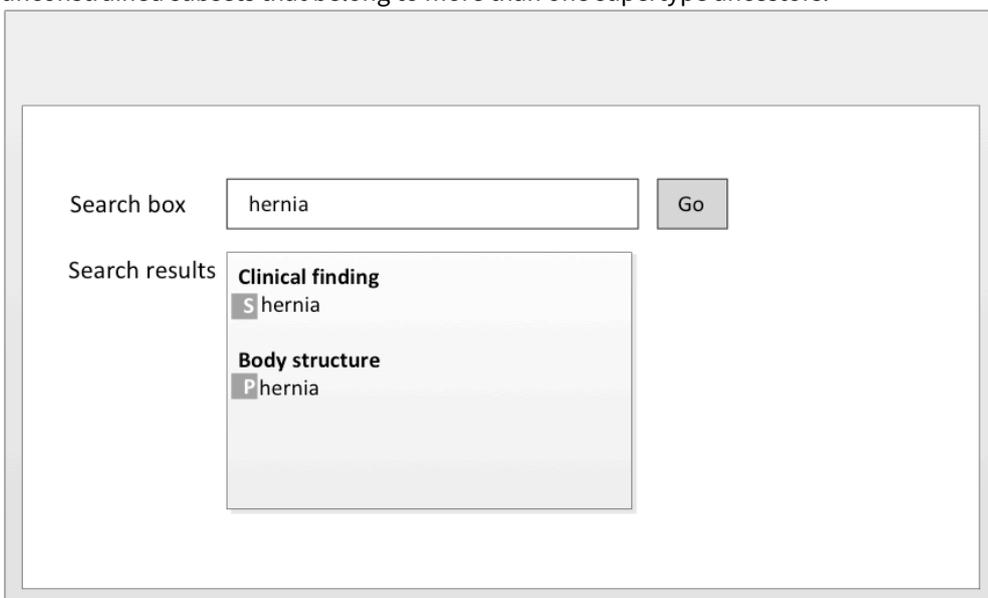
## 5.2. Distinguish Identical Terms of Different Concepts

A search for a term (e.g. "fundus") may return multiple identical matches. Fully specified name Descriptions are always unique at a specific release. However, the preferred terms or synonyms may not always be unique, as illustrated in [Figure 5.2-1](#).



**Figure 5.2-1: Hovering over identical terms of different Concepts**

Identical terms of different Concepts can be distinguished by displaying the search results under different headings according to which supertype ancestors they belong to. This technique is particularly useful for searching large unconstrained subsets that belong to more than one supertype ancestors.

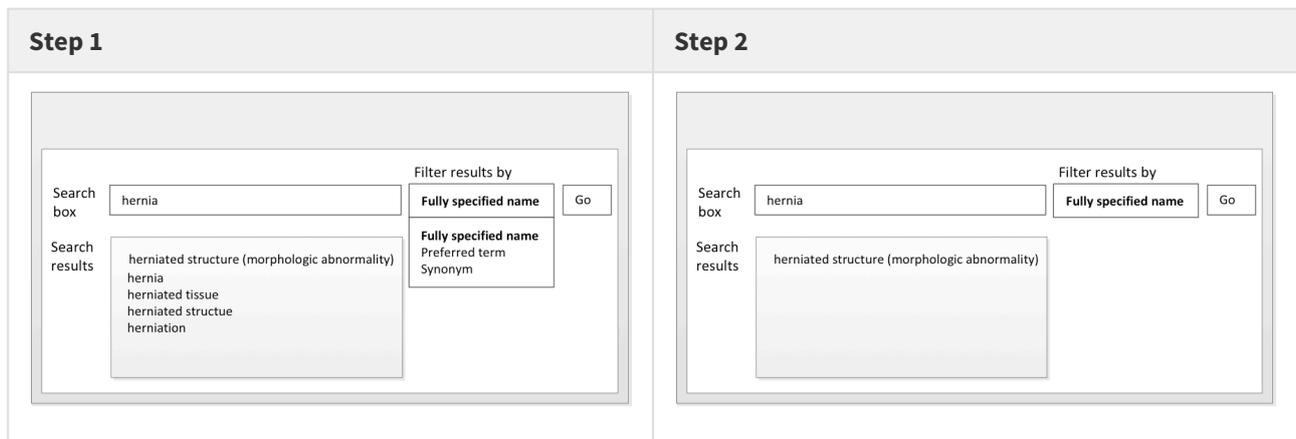


**Figure 5.2-2: Categorizing search results under subtype hierarchies**

### 5.3. Avoid Multiple Hits on the Same Concept

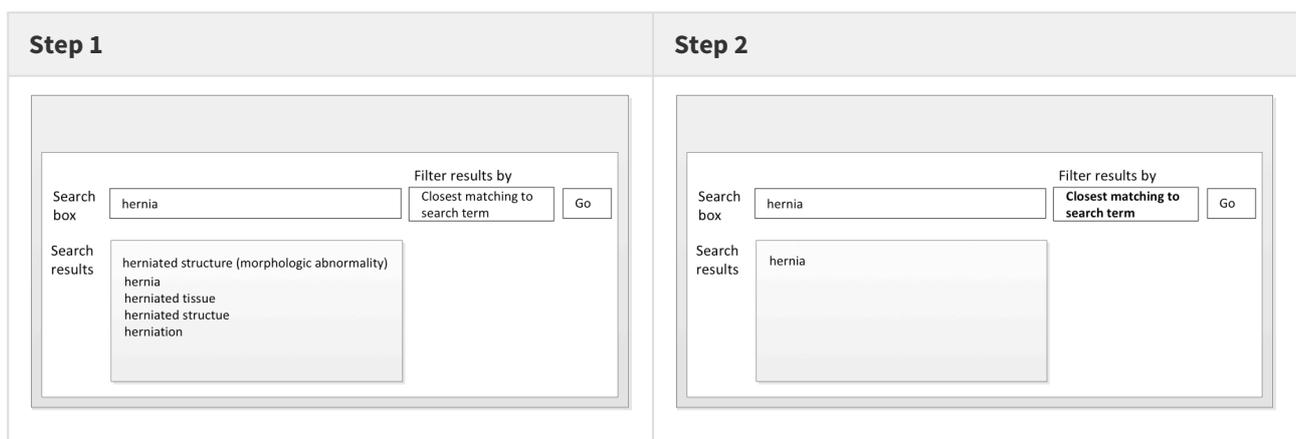
In many instances, several synonyms associated with the same Concept contain the same keyword. For example, | herniated structure |, | hernia |, | herniated tissue |, | herniated structure | and | herniation | all begin with "hernia". A search for the target Concept "hernia" would return the first phrase found during the search. Designing a filter that gives the user the option of filtering the results by description type such as the Preferred Term or the Fully Specified Name can significantly reduce the results, as illustrated in [Figure 5.3-1](#).

#### Avoid Multiple Hits on the Same Concept by Filtering the Search by Description Type



**Figure 5.3-1: Filtering search results to show only the Fully Specified Name (search also constrained by supertype ancestor)**

Designing a filter gives the user the option of filtering the results to show only the Description associated with the Concept that is the closest match to the search term, as illustrated in [Figure 5.3-2](#)



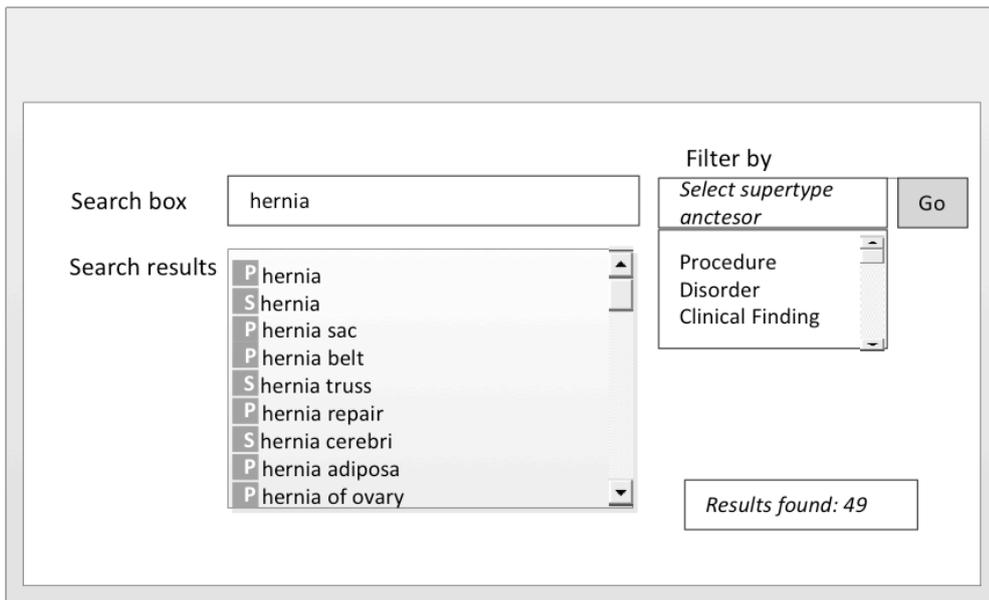
**Figure 5.3-2: Filtering search results to show only the closest match to the search term, hernia (search also constrained by supertype ancestor)**

## 5.4. Rationalize Search Results by Subsumption Checking

Filtering a search by using subsumption checking is an effective technique to enhance the display of search results. It reduces the list of search results by nesting the subsumed Concepts under the general Concept. If the user wishes to select a narrower Concept, they can expand the node to select the nested subtype.

### An Example Illustrating Rationalization by Subsumption Cross-Checking

Unconstrained search for descriptions that begin with "hernia" returns a total of 49 Concepts which belong to many subtype hierarchies such as [Clinical Finding](#) and [Morphologic abnormalities](#).



**Figure 5.4-1: Rationalize search results by subsumption checking**

Constraining a search for "hernia" by the "clinical findings" supertype returns 35 matches which is still considered to be a long list, many of which are subsumed by the general Concept "hernia". [see [Constrain searches by supertype ancestors](#)]

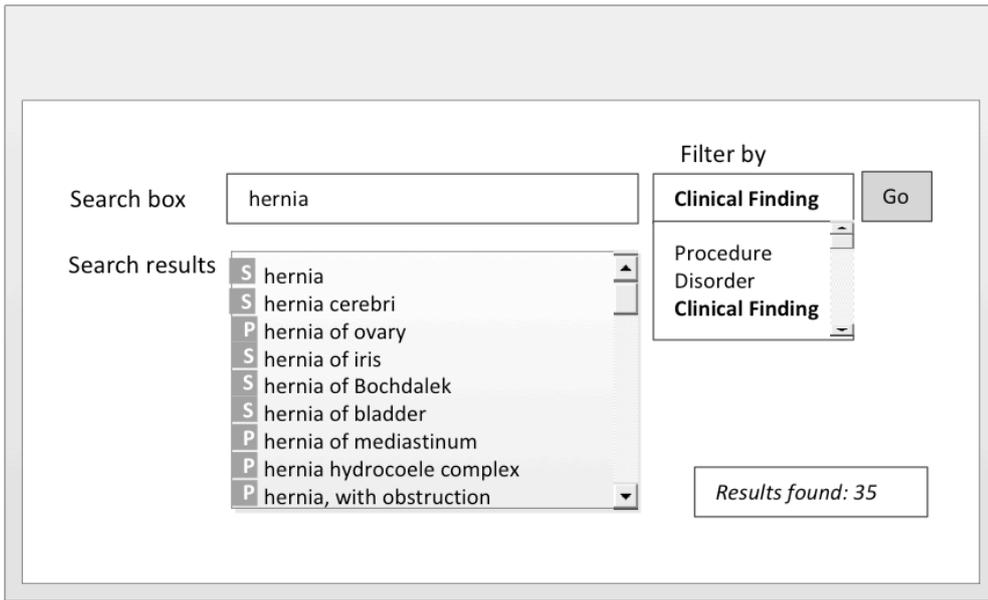


Figure 5.4-2: Rationalize search results by subsumption checking - clinical findings only

Rationalisation by subsumption cross-checking can further reduce the matches by nesting the subsumed Concepts such as | perineal hernia | under | hernia |. If the user wishes to select a narrower Concept, they can expand the node to select the nested subtype.

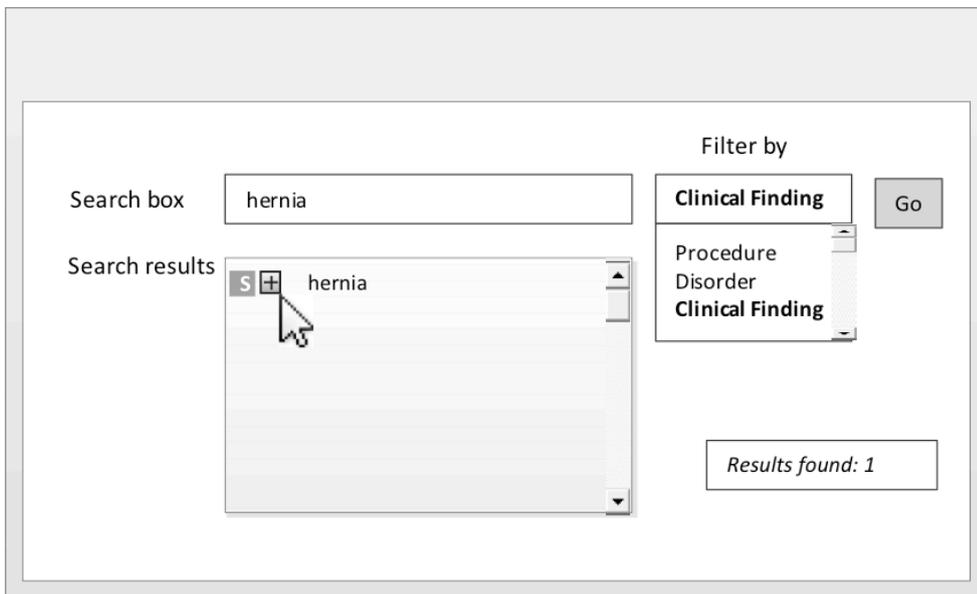


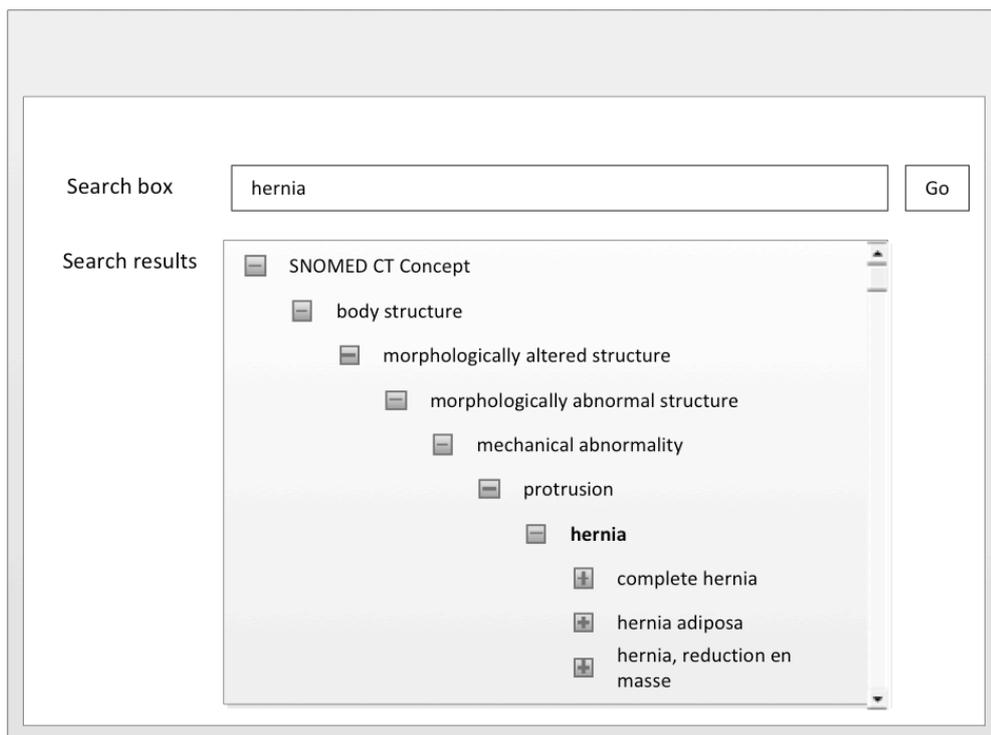
Figure 5.4-3: Subsumed Concepts nested under "hernia"

## 5.5. Display Navigation Results Effectively

Results navigation is useful feature for search browsers to have in addition to the conventional display of search results (see [5.1. Order Search Results Rationally](#) and [5.4. Rationalize Search Results by Subsumption Checking](#)). The decision to design navigation results in browser may depend on space availability in the application's user interface. This sub-section describes two different types of hierarchies for browsing search results.

### 5.5.1. Using the Subtype Hierarchy

The most visible hierarchical construct in SNOMED CT is the subtype hierarchy. This is constructed using a set of logical rules.



**Figure 5.5.1-1: Matches shown in the subtype hierarchy**

The primary use of the SNOMED CT subtype hierarchy is to support effective retrieval and aggregation of data.

#### Example:

The Concept "Laparoscopic emergency appendectomy" can be reliably located by subtype navigation from any of its supertypes: "appendectomy," "laparoscopic appendectomy" or "emergency appendectomy."

#### Enhanced Hierarchical Displays

It is possible to start at the top of hierarchy and navigate from parent to child in order to find a Concept or term in SNOMED CT. A more efficient approach, however, is to use the hierarchy to supplement a keyword search by enabling the user to look at related Concepts in order to consider them as alternative matches, or to check the context of a search result. The following examples illustrate these two uses of the SNOMED CT hierarchy.

## Examples:

### 1. Checking supertypes:

A user wishes to find a *Description* that relates to the condition of a patient who is hypersensitive to an allergen. The user performs a search on the *keyword* "Hypersensitivity" and finds an exact match. Before the user selects the *Description* for inclusion in the patient record, they check the *Fully Specified Name*, which is | Sensitivity (finding) |. The user then checks the *hierarchy* and discovers that the selected *Concept* has | Psychological finding | as an *ancestor*, which indicates that this is not the correct *Description* to use in this context.

### 2. Checking subtypes:

A user wishes to find a *Description* that relates to the condition of a patient who is hypersensitive to an allergen. The user searches for the *keyword* "allergy," and finds one *Concept* having a *Description* that is an exact match. The user then looks at the *children* of the *Concept* (i.e. those *Concepts* immediately below it in the *hierarchy*). One of the *children* has the preferred *Description* | Contact Hypersensitivity | which matches the user's intended meaning. The user selects this *Concept* for inclusion in the patient record.

## 5.5.2. The Challenge of Subtype Navigation

The same hierarchy can be used for data entry navigation following search but it is not designed for this purpose. Its depth and breadth are determined by logical rules of subsumption rather than by usability. As a result: There is no upper limit on the number of subtypes a *Concept* may have. This is true because there is no rule that determines the number of subtypes that a real world *Concept* may have. However, long lists of options are not conducive to effective data entry.

There is no fixed limit to the number of hierarchical steps between a generalized *Concept* and its most refined subtype. This is true since there is no preordained limit on the extent of possible refinement of a real world *Concept*. However, data entry procedures that involve stepping through several levels of choices before reaching the required selection impair usability.

The subtypes of a *Concept* do not have any particular order. The | is a | Relationship is primarily a property of the subtype *Concept* and does not express an ordinal position. This is true because logical subtypes are inherently an unordered set. However, a user is likely to find it easier to locate their required selection if members of hierarchical lists are displayed in some recognizable order.

The issues of depth, length and order noted above are also subject to change between releases. The addition of an intermediate *Concept* or reclassification after the addition of new defining characteristics will introduce new layers in the hierarchy. Some *Concepts* will then move from the list of immediate subtypes of a *Concept* to become subtypes of a more refined *Concept*. Hierarchical changes may sometimes simplify navigation by reducing the number of choices at a given hierarchical level. However, the general effect of improvements in the subtype hierarchy will be to increase its depth and thus to increase the number of steps from a particular general *Concept* to its most refined subtypes.

The poly-hierarchical structure allowing for a concept to have more than a single parent concept means that there may be many routes from a given *Concept* to its more general ancestors. This means that some of the choices presented for user selection are redundant since they simply offer alternative routes to the same *Concept*.

Routine use of subtype hierarchy navigation is not recommended for data entry. However, despite the drawbacks listed above, the subtype hierarchy may be useful for undertaking an exhaustive search for a particular refined *Concept*.

### 5.5.3. Using the Navigation Hierarchy

*SNOMED CT subtype Relationships* provide a logical semantic *hierarchy*. Often it is possible to view parts of the terminology and select particular *Concepts* by navigating through this *subtype hierarchy*. However, there are many situations in which the pure *subtype hierarchy* does not provide an ideal route for navigating the *hierarchy* (see [5.5.1. Using the Subtype Hierarchy](#)).

Navigation hierarchies can be used to drive some types of structured data entry. Navigation hierarchies can order data in sensible ways by priority, or by some readily understood convention (e.g. "cranial nerve order" or "pharmacy products in the order of strength"). Navigation hierarchies can be used for diverse purposes (e.g. "topics related to diabetes").

*Navigation links* are used to provide an alternative route through parts of the terminology. A *navigation link* can link any two *Concepts* together to identify a useful route for *navigation*. Each of the *navigation links* is directional, linking a navigational parent *Concept* to a more refined navigational *child Concept*. However, unlike the *subtype relationship* the presence or absence of a *navigation link* neither adds to nor subtracts from the definition of either of the *Concepts* that it links.

Some *Concepts* may exist only to provide nodes in a *navigation hierarchy*. These *Concepts* are *subtypes* of | Navigational *Concept*/ and play no part in the semantic definitions of any other *Concept*.

Navigational hierarchies are represented as Reference Sets, which are available in the International Release of SNOMED CT or they can be created from scratch to meet specific user needs. Navigational hierarchies created from scratch do not have to represent subtypes or logical relationships unless it is required. (see [6.2.5 Implementing Navigation Hierarchies](#)).

- common viral disease
  - common cold
  - herpes simplex
  - Human immunodeficiency virus infection
  - Infectious mononucleosis
  - influenza
  - measles
  - mumps
  - rubella
  - varicella-zoster virus infection
  - verruca vulgaris
  - viral gastroenteritis
  - viral hepatitis
  - viral meningitis
  - viral pharyngitis
  - viral respiratory infection
  - viral disease (all)

**Figure 5.5.3-1: An example of a handcrafted navigation hierarchy showing a list of common viral diseases**

## 5.6. Use Mnemonics and Personal Favorites for Data Entry

Groups of people, such as practitioners of a discipline or specialty, frequently use similar sets of Descriptions and Concepts. Lists of widely understood (or easily learned) abbreviations or mnemonics that allow rapid entry of these commonly used Concepts are recommended as a way of accelerating repetitive recording.

A similar facility may also be useful for individual users or organizations that have sets of Descriptions and Concepts that they use frequently. Providing an easy way for users to store personal favorites following a search and recall these favorites with user-defined abbreviated access terms will enhance usability and significantly increase the speed of data entry.

User guidance may be necessary to minimize the risk of overusing the shortcuts. Unless the general search facilities are also easy to use, it is likely that users will favor the shortcuts even when it would be more appropriate to use a more accurate but less accessible Concept. An unchecked bias toward easy to record Concepts may lead to deterioration in data quality, statistical anomalies, and in the worst case, inappropriate treatment.

## 6. Data Entry

Data entry is the process by which a user submits information containing relevant SNOMED CT Concept identifiers for storage in a record system (e.g. an EHR system).

The way data entry is carried out depends on the setting in which it is performed. In one simple approach a user selects a term from a list of search results, confirms the selection and the Identifier or Description of the selected Concept is added to the relevant field in a database record. This guide considers a range of different modes of data entry which can make it quicker and easier to record different types of information in particular situations.

The focus of this section is on the non-search aspect of data entry but it is also important to recognize that this is usually only one part of the process of recording clinical information. Therefore, the way SNOMED CT data entry is integrated with the overall process of data capture also needs to be considered.

### 6.1. Benefits of Using SNOMED CT for Data Entry

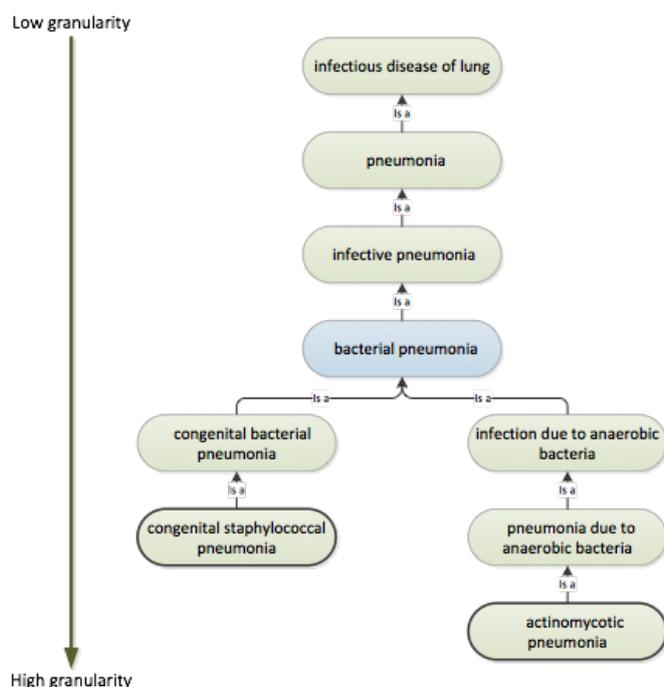
#### 6.1.1. Specificity and Generalization

SNOMED CT hierarchies are subtype hierarchies (made up of 'is a' relationships) that allow for clinical information to be entered at a level of detail sufficient for the specific clinical situation or possible at the given point in time.

For example, it is possible to specify a clinical finding at very high level of detail, which is relevant and important for clinical treatment. The 'is a' relationships allow for SNOMED CT hierarchies to be automatically traversed, and to enable terminological reasoning of common sub- and supertypes. This means that Concepts can be automatically aggregated into more general types, which is useful for management purposes and for statistics.

#### Example:

Traversing the supertype hierarchies of the two concepts | actinomycotic pneumonia | and | congenital bacterial pneumonia | will show that both Concepts are types of | bacterial pneumonia |.

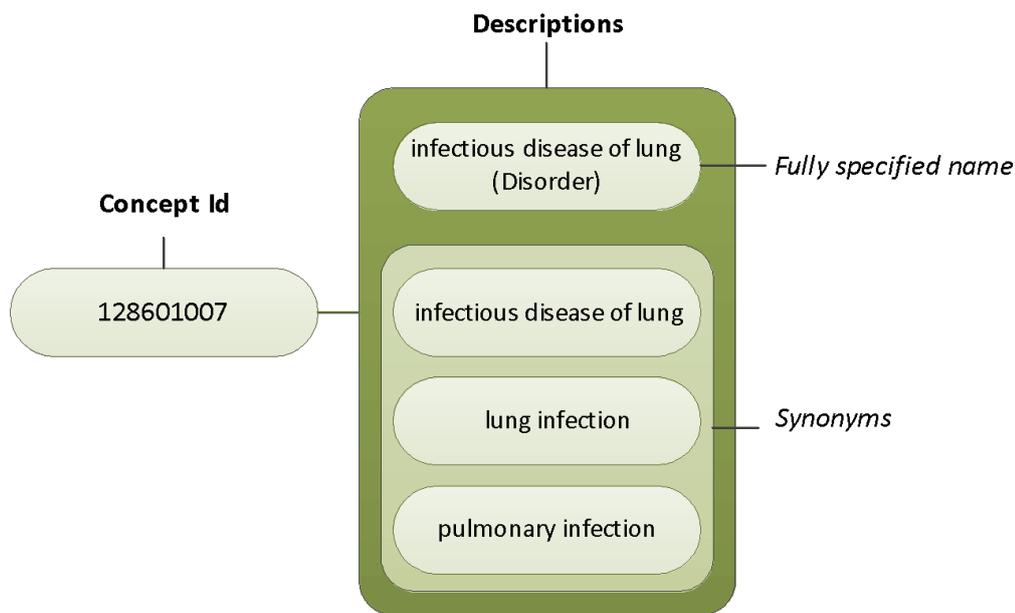


**Figure 6.1.1-1: The subtype hierarchies link specific detailed concepts to general concept with less granularity**

## 6.1.2. Clinical Language and Coding

The unique Concept identifiers of SNOMED CT Concepts are numeric identifiers and do not contain information related to the meaning of a Concept or Description. This means it is not possible to infer anything about the meaning of a Concept from the numeric value of the Identifier or from the sequence of digits that form the identifier.

The meaning of a Concept can be determined from relationships to other Concepts and from associated Descriptions that include human readable terms. In the context of data entry, this means that entering SNOMED CT Concepts must be done by using associated Concept Descriptions, and hence by use of the clinical spoken language instead of using the Concept codes for data entry.



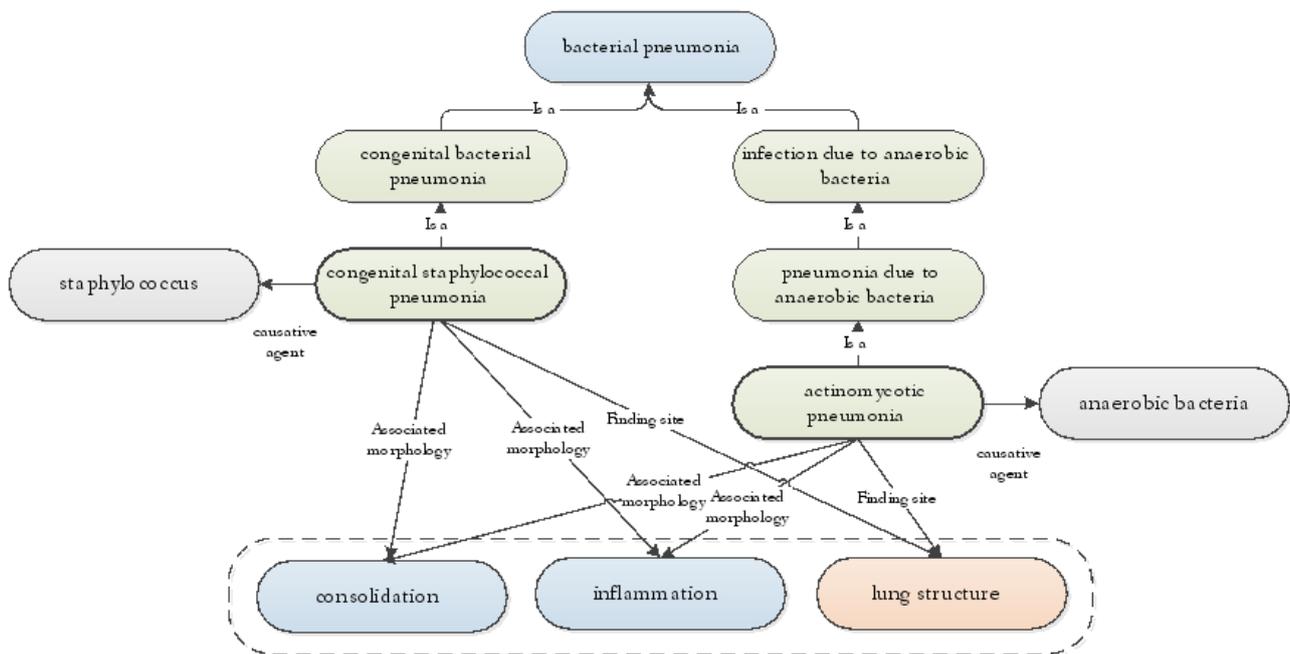
**Figure 6.1.2-1: SNOMED CT Descriptions support the use of clinical language for documentation and the synonyms support flexibility in the terms used to represent the Concepts**

Additionally, SNOMED CT allows for coding different Descriptions that belong to the same Concepts. This support flexible data entry, by enabling the possible to use the Descriptions preferred in a specific organization or within a specific user group.

## 6.1.3. Data Entry as Foundation for Meaning-Based Retrieval

Use of SNOMED CT for data entry has implications for how clinical data can be used and retrieved. The SNOMED CT relationships support meaning-based retrieval of clinical data, which can be useful for creating specific clinical overviews or for aggregating data with specific characteristics (meaning).

The figure below shows some of the defined relationships that are included in SNOMED CT to define the two Concepts | actinomycotic pneumonia | and | congenital bacterial pneumonia |. The Concepts have different causative agents; however the associated morphology and the finding site of the findings are equivalent. This means, that both Concepts would be part of a result of queries that retrieve clinical findings with the finding site | lung structure | and the associated morphology | inflammation | or | consolidation |.



**Figure 6.1.3-1: Meaning based retrieval**

Such types of meaning-based queries are possible to develop because of the integrated clinical meaning which is made explicit by the defined relationships. Terminological reasoning therefore makes it possible to develop queries that retrieve all Concepts with a specific meaning. In many cases this approach is much more effective and efficient than manually specifying which Concepts to retrieve based on the unique Concept Id's.

### 6.1.4. Effective and Consistent Data Entry

SNOMED CT can support consistent data entry. The unique Concept Identifiers support unambiguous representation of the clinical content applied in Clinical Information Systems.

However, the flexibility of SNOMED CT allows entering expressions in various ways which may hamper consistency and retrieval. Structured data entry mechanisms such as, having a pre-defined set of Concepts for a specific clinical data entry field in a form can constrain the number of Concepts visible to documenting clinicians and enhance consistent use of SNOMED CT Concepts. [see Examples of applied mechanisms of structured data entry]

Therefore it is important that each SNOMED CT-implementation includes strategies for how to balance the flexibility of SNOMED CT when entering data into the clinical record.

## 6.2. Structured Data Entry

Structured data entry can be applied to support effective data entry and ensure consistent use of Concepts. When using structured data entry methods, SNOMED CT encoded data can be selected in a variety of ways. Some of these involve direct selection of terms while other ways involve encoding which may result from responses to simple choices or entry of particular data values. The sub-section below shows requirements for structured data entry and examples of mechanisms for data entry.

### 6.2.1. Requirements for Structured Data Entry

Data entry needs to be easy in terms both of the time and effort required to record the information needed, but also easy in the way it fits in with working practices.

User interfaces should make it easy to enter data that includes SNOMED CT where this is required to enable effective retrieval and reuse. Hence, data capture is only worthwhile if the data captured can be usefully reused, and therefore data capture needs to be designed to meet requirements for subsequent retrieval.

SNOMED CT encoded data should be rendered in ways that are easily read and understood as part of the clinical record. Approaches to data entry therefore need to be tailored to the way different groups of clinicians work and think. On the one hand, the application of SNOMED CT must be done by a common approach across the various user interfaces to support consistency. On the other hand, the system should not be designed as one size fits all, as this will decrease the applicability of the system. SNOMED CT supports flexible and user-oriented data entry mechanisms, allowing for detailed and very specific data entry as well as bridging different levels of granularities applicable for retrieval of common patterns.

Applying SNOMED CT in EHR has implications for how data can be utilized for effective clinical documentation, enhanced retrieval and decision support functionalities. To support effective display and analysis of SNOMED CT coded data, questions about an individual patient or a selected population must have answers that are:

- Accurate – no false negatives
- Precise – no false positives
- Timely – available when needed
- Efficient – without disproportionate cost, time and effort
- Consistent – independent of how the data was entered

To meet the above requirements, collected data must be stored in a consistent form that enables reliable retrieval.

### 6.2.2. Balance Interface Design and Concept Selection

The way information is captured in the clinical record can differ between systems. The choice of Concepts and their representation in the record has implication for interpretation and retrieval.

Requirements for data retrieval are an important determinant of requirements for data content and representation, and the following two overall guidelines can be used to govern the Concept selection and representation:

- To support effective reuse a health record must make it possible to answer relevant questions accurately and efficiently;
- To answer relevant questions information must be selectively retrieved so it can be displayed or analysed.

As the following examples illustrate, the same information (clinical meaning) may be captured in different ways. If information is represented according to the way it is captured it may be difficult to answer simple questions.

*A question such as 'does the patient have a family history of diabetes mellitus?' can expand to:*

1. Do they have a family history form in which 'diabetes mellitus' is checked as present?
2. Do they have a family history record in which the code for 'FH diabetes mellitus' is present?
3. Do they have text that is tagged with the code for 'diabetes mellitus' in the context of a section of text tagged as 'family history'?
4. ... n. *There are also many other possible data capture representations to consider*

### 6.2.3. Default Context

Concepts in SNOMED CT carry a "soft-default" context, which means that, unless stated otherwise within the Description or the definition of the Concept, the term should be interpreted as follows:

For a clinical finding that:

- The finding has actually occurred (vs. being absent or not found).
- It is occurring to the subject of the record (the patient).
- It is occurring currently or at a stated past time.

For a procedure that:

- The procedure was completed.
- It was performed on the subject of the record (the patient).
- It was done in the present time or at a stated past time.

The only exceptions are Concepts whose Description actually contains a specific context (e.g. father smokes), and these are all grouped in a special hierarchy (situation with explicit context).

When recording in the patient record, post-coordinated expressions or free text should only be used to add additional detail to the coded item. They should never negate or modify the meaning of the coded item as this would fundamentally changes its meaning.

For example, free text phrases such as 'family history of', 'absent', 'not found' should not be applied to SNOMED CT coded data.

### 6.2.4. Examples of Interface Design AND Concept Selection

#### Implicit and Explicit Meaning of Interface Terms

The illustrations below shows a template designed for the purpose of documenting diseases that occur in the patients family history.

The first template below shows a heading that specifies the context of the template, namely documentation of 'family history', whereas the checkboxes allow for specifying the family history of the disease. The direct (lexical) map between the interface term "Diabetes Mellitus" without considering the actual context will be the clinical finding Concept | Diabetes Mellitus (Disorder) |.

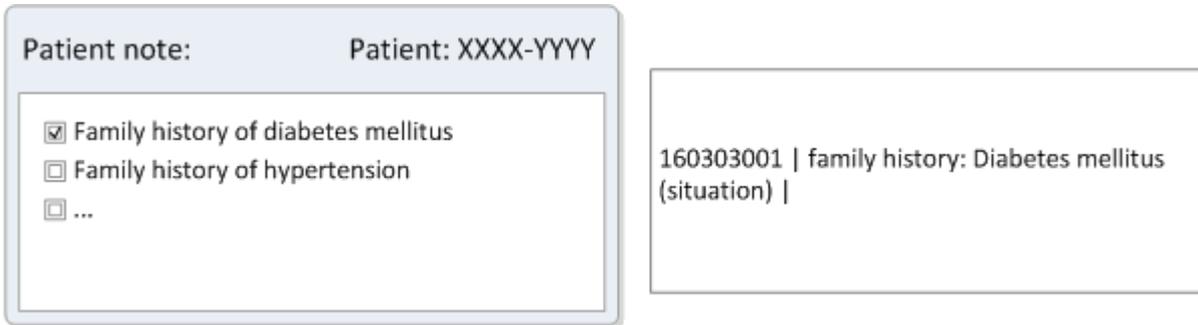


The figure illustrates a user interface for documenting family history. On the left, a form titled "Family history" for "Patient: XXXX-YYYY" contains a list of diseases with checkboxes: "Diabetes Mellitus" (checked), "Hypertension", and "...". On the right, a separate box displays two SNOMED CT codes: "160265008 | family history taken (situation) |" and "73211009 | diabetes mellitus (disorder) |".

**Figure 6.2.4-1: Family history form with interface term of implicit meaning**

However, the appropriate Concept for data entry is the situation Concept | family history : Diabetes mellitus (situation) | based on the meaning of the interface term.

The form below shows a patient note with check boxes to enter diseases that occur in the patient's family history. Here, there is a direct (lexical) map between the interface term "family history : Diabetes mellitus" and the meaning of the interface term, | family history : Diabetes mellitus (situation) |.



Patient note: Patient: XXXX-YYYY

Family history of diabetes mellitus  
 Family history of hypertension  
 ...

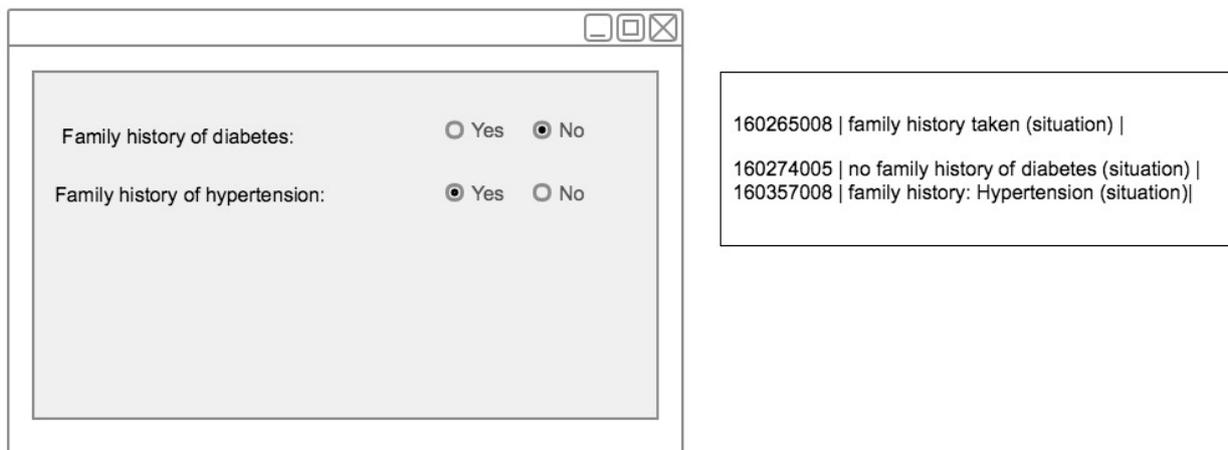
160303001 | family history: Diabetes mellitus (situation) |

Figure 6.2.4-2: Patient note form with interface term of explicit meaning

It is therefore a requirement that careful thought be given about the actual meaning of each entry when selecting Concepts to represent a data entry, regardless of the specificity (or correctness) of the interface term.

### Entry of Presence and Absence

The example below shows another typical way of documenting occurrences or negations of specific findings. The interface terms 'yes' and 'no' can be mapped to the qualifier values | yes (qualifier value) | and | no (qualifier value) |. However, these Concepts are meaningless without knowing the finding or situation to which they relate. The appropriate mapping for these entries would therefore be expressions that specify the presence or absence of the actual finding, situation, etc. In this example the Concepts that represent respectively 'Yes' and 'No' would be the situations | no family history diabetes (situation) | and | family history: Diabetes mellitus (situation) |.



Family history of diabetes:  Yes  No  
Family history of hypertension:  Yes  No

160265008 | family history taken (situation) |  
160274005 | no family history of diabetes (situation) |  
160357008 | family history: Hypertension (situation)|

Figure 6.2.4-3: Entry of presence and absence

## 6.2.5. Examples of Applied Mechanisms of Structured Data Entry

### Short List of Options

User selects from a small list of possible [Descriptions](#) applicable to a particular field in a template or step in a protocol. A Simple [Reference Set](#) with corresponding [Descriptions](#) from a particular Language [Reference Set](#) may specify the set of applicable [Descriptions](#).

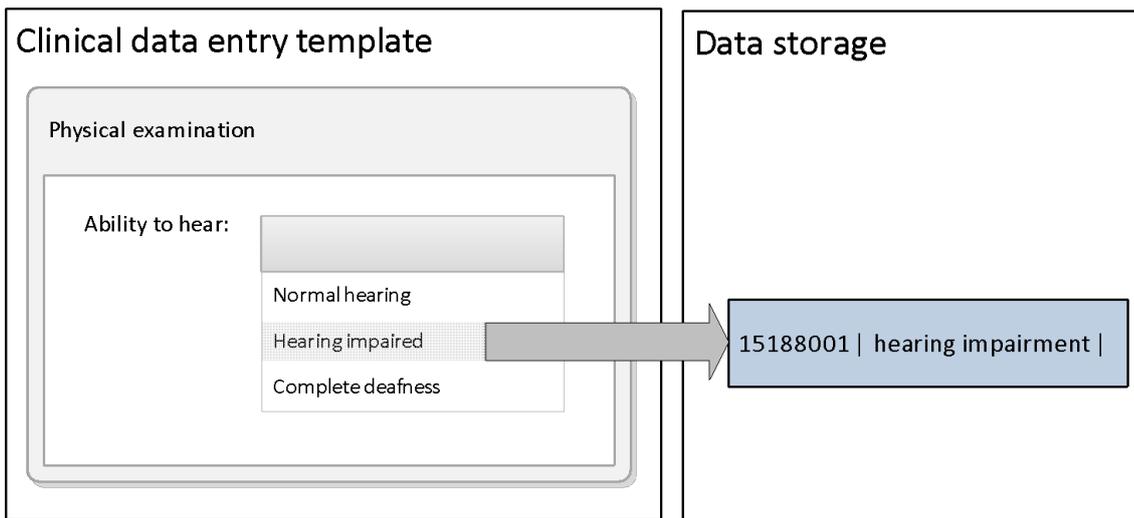


Figure 6.2.5-1: Data entry using a simple reference set

### Longer Lists Require Effective Search Techniques

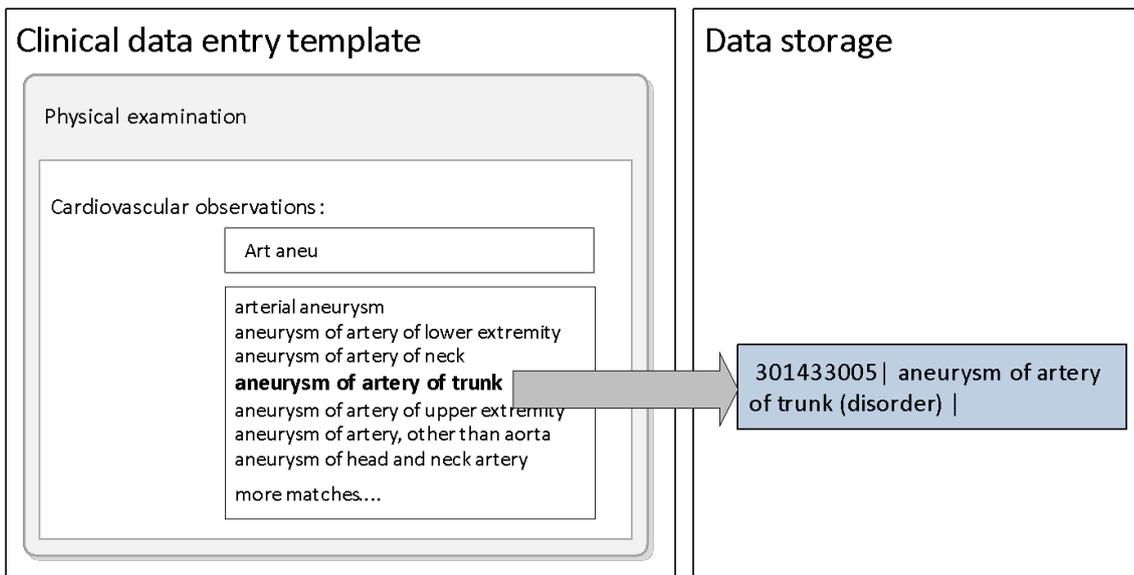


Figure 6.2.5-2: Data entry using a large Reference set requires use of search techniques

See section 4. [Optimizing Searches](#), for further information about effective search techniques.

### Binding a Concept to a Data Control Used for Entering a Numeric or Other Value

When a value is entered in this control it is labeled with the appropriate Concept Identifier and added to the record.

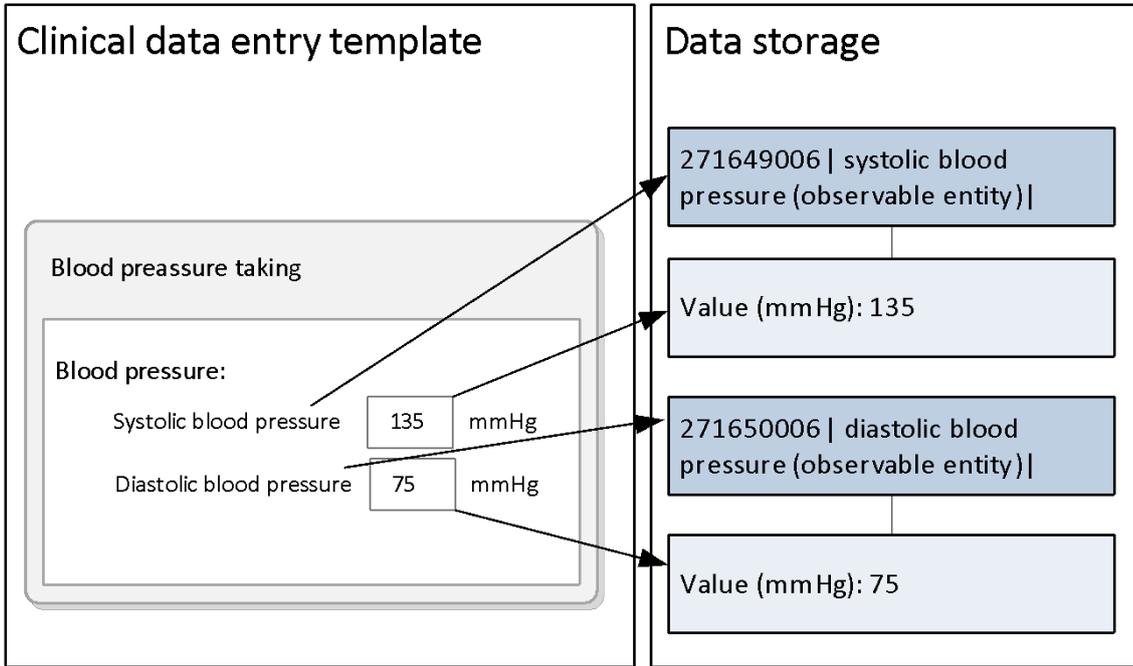


Figure 6.2.5-3: Data entry of numeric values

### Association of a Concept with Particular Options Presented by a Check Box, Option Button or Other Data Entry Control

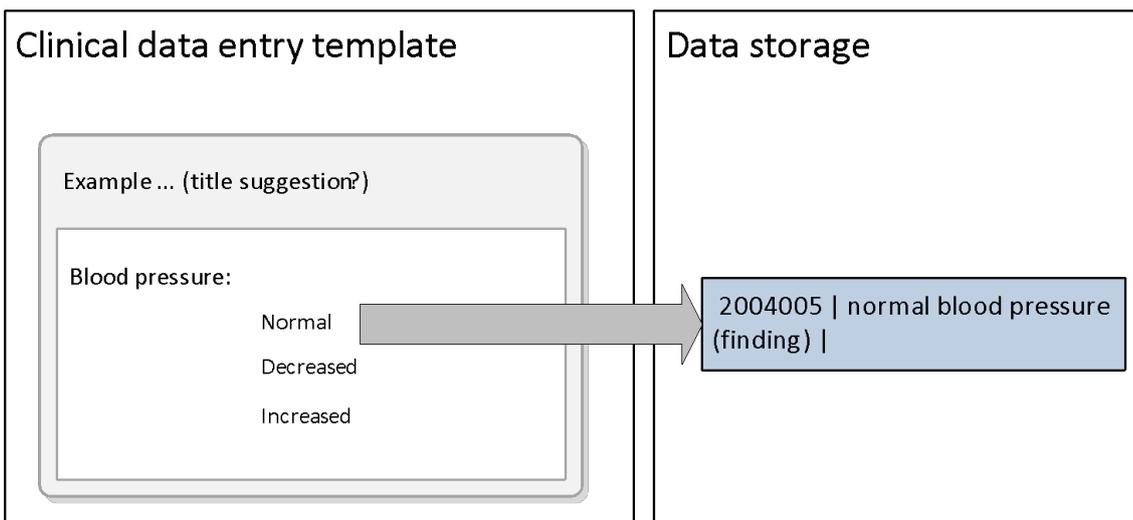
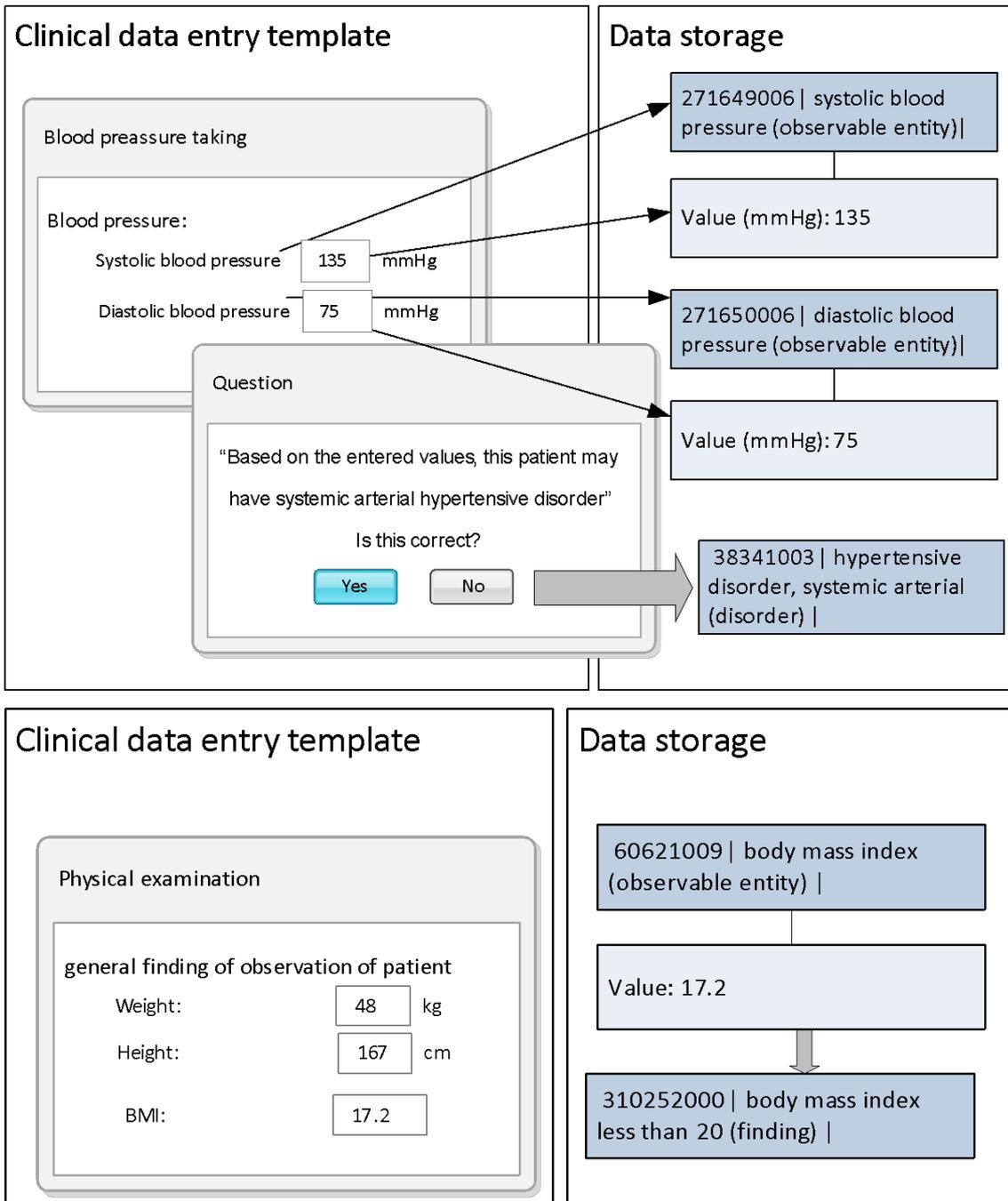


Figure 6.2.5-4: Data entry using radio buttons / check lists

When selections are made using this control the appropriate Concept identifier is added to the record.

### Binding Concept to a Particular Combination of Values or the Result of a Computation Involving Several Items of Previously Entered Data



**Figure 6.2.5-5: Two examples of how data entry of specific values can be foundation for automatic addition of appropriate SNOMED CT clinical findings**

When specific values are entered the appropriate (defined) Concept identifier is added to the record. The above examples show how specific numeric values can lead to specific clinical findings, based on pre-defined rules.

The example at the bottom of Figure 39, illustrate that the BMI can be automatically calculated by the entered values for body weight and height. Moreover, it is possible to develop rule-based coding (i.e. algorithms that assign BMI values between 20 and 24 to the SNOMED CT Concept 35425004 | normal body mass index |, or to a more specific Concept 412768003 | body mass index 20-24 - normal |.

Value	Concept
< 20	<ul style="list-style-type: none"> <li>• 6497000   decreased body mass index  </li> <li>• 427090001   body mass index less than 16.5  </li> <li>• 310252000   body mass index less than 20  </li> </ul>
25 <	<ul style="list-style-type: none"> <li>• 48499001   increased body mass index  </li> <li>• 162863004   body mass index 25-29 - overweight  </li> <li>• 162864005   body mass index 30+ - obesity  </li> <li>• 408512008   body mass index 40+ - severely obese  </li> </ul>
20 -24	<ul style="list-style-type: none"> <li>• 35425004   normal body mass index  </li> <li>• 412768003   body mass index 20-24 - normal  </li> </ul>

## 6.3. Constraining Data Entry

Constraining searches refers to disabling Concepts or Descriptions in searches and navigation or allowing these search and navigation results to be filtered. [See [Constrained Searches](#)] However, some Concepts and Descriptions that are suitable to be displayed as search results or navigation results, may not be suitable for recording in a patient record once search constraints have been applied. Constraining data entry refers to applying constraints so that only Concepts and Descriptions suitable for recording are selectable.

### 6.3.1. Constraining Data Entry Based on Context

All fields (data elements) used for data entry must be analyzed to understand what underlying context is implied. The appropriate [Concepts](#) should then be selected for the [value set](#) of each field. [Concepts](#) from the Clinical Findings, Procedures, and Observable entities hierarchies can be used directly if the default assumptions are true. Otherwise, Concepts from the Context Dependent hierarchy should be selected to explicitly code the underlying context of the clinical idea.

Particular care must be taken with systems that enable real time post-coordinated constructs to ensure that the appropriate context attributes are included.

To pre-coordinate Concepts that do not already exist in SNOMED CT, care must also be taken to determine if any axis modification is shifting the meaning of a Concept. If so, it should be moved to the *Situation with explicit context* hierarchy.

### 6.3.2. Constrain Data Entry by Concept Description Status

Constraining data entry by Concept Status and Description Status should be done in order to avoid recording Concepts and Descriptions which for some reason have been inactivated. Inactive Concepts or Descriptions have been assessed to be unsuitable for recording. Therefore they should not be included as a possible result of a search, which would make them candidates for clinical recording.

### 6.3.3. Constrain Data Entry According to Supertype Relevance

Constraining data entry according to supertype relevance is a technique that can support data entry, in which only particular sub-hierarchies that include Concepts relevant for the specific data entry point can be browsed or displayed.

**Example:**

If a pick list in a data entry template should allow the selection, a specific "cardiopulmonary bypass operation", the SNOMED CT subtype hierarchy to specify what Concepts are to be included in the list, namely the descendants of the Concept "63697000 | cardiopulmonary bypass operation | ". Constraining data entry according to subtype relevance can be done using the transitive closure table, as shown in the figure below. The transitive closure table contains relationships to all supertypes ancestors of each defined Concept. A query for the set of Concepts, where the Concept "63697000 | cardiopulmonary bypass operation | "(shown as Concept B in Figure 40) is a supertype, will produce the constrained set of Concepts relevant for creating specific pick lists.

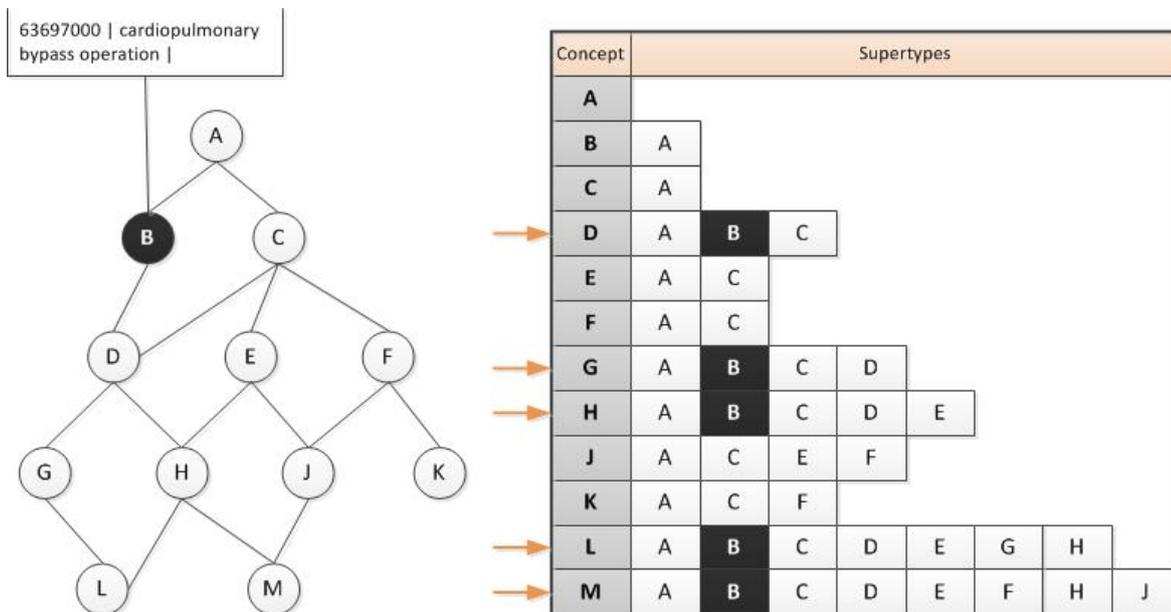


Figure 6.3.3-1: Using transitive closure to constrain data entry according to supertype relevance

Identifying selected portions of the SNOMED CT hierarchy may not be a sufficient constraint for entering data into a record. Constraining data entry through the use of Reference Sets may be more sufficient for producing the required set of Concepts relevant for data entry. [see 6.4.4 Constrain data entry using Reference Sets]

### 6.3.4. Constrain Data Entry Using Reference Sets

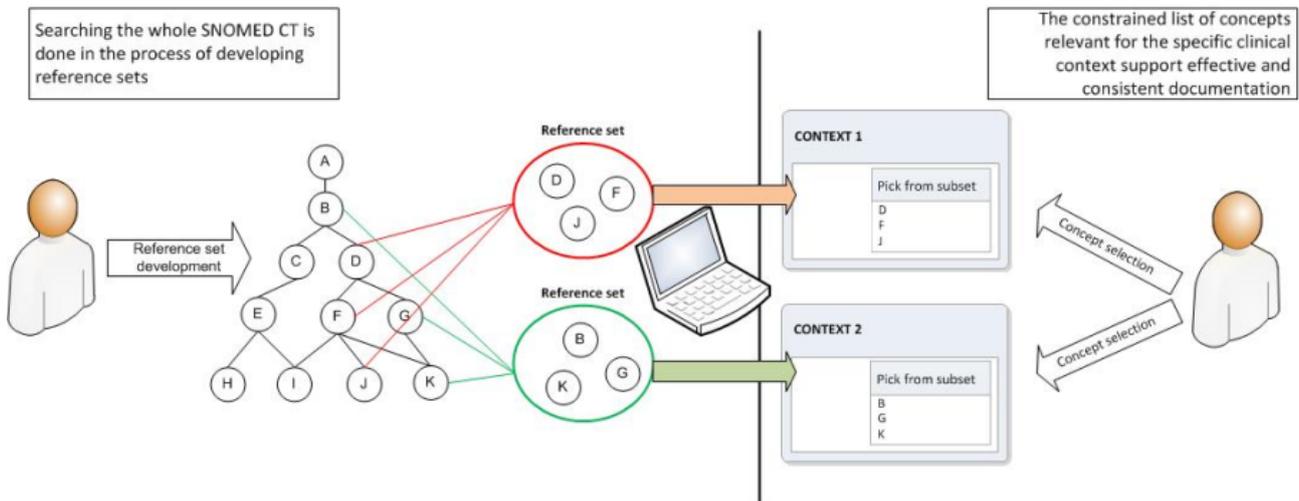


Figure 6.3.4-1: Constraining data entry using Reference Sets

In some cases, identifying selected portions of the SNOMED CT hierarchy may be a sufficient constraint for entering data into a record. However, it is not always sufficient if Concepts from multiple hierarchies are required, or if there is a need to hone the entry options from the full hierarchy. To meet these requirements applications should allow data entry to be constrained by Reference Sets.

Applications should be able to:

- Permit or prevent the entry of Concepts or Descriptions that are members of a specified Simple Reference Set.
- Permit entry of Concepts or Descriptions that are ordered in priority using the Ordered Reference Set.

**Example:**

A UK GP practice implemented a clinical application with a data entry mechanism that:

- Prevents the entry of Concepts in a Simple Reference Set that contains all Concepts that are in the Non-Human Simple Reference Set so human Concepts can only be entered and stored in the clinical record;
- Enables the entry of Concepts in the "UK Administrative Reference Set" only when entering information in an administrative context.
- Enables the entry of Concepts in search results that is prioritized according to the order of disease prevalence using an Ordered Reference Set. Descriptions of Concepts associated with high disease prevalence are displayed before those with lower priority.

**Example:**

A specialty system might prompt for confirmation when the user records a procedure not in a specified specialty Simple Reference Set.

## 6.4. Entering Refinements for Post-Coordinated Expressions

SNOMED CT contains **Concepts** that allow many clinical ideas to be represented by a single *Concept identifier*. However, SNOMED CT also allows more specific ideas to be represented even when they cannot be expressed by a single Concept. This is achieved by adding **refinement**s to a selected **Concept** creating a **post-coordinated** expression containing several Concept Identifiers.

Several types of post-coordinated data are outlined in this section from the perspective of data entry. These include refinement, qualification and combination. The requirements for and relevance of each of these will depend on decisions about data representation within patient records.

### 6.4.1. Entry of Refined Defining Characteristics

The application may allow a user to refine a Concept by selecting a subtype of one of its defining characteristics. A defining characteristic is a relationship to a target Concept that is always necessarily true from any instance of the source Concept.

Refinement options may be entered by selecting from hierarchical lists showing subtype values. Simple lists or option buttons could support selection from limited sets of possible refinements. Wider ranges of possible refinement could be facilitated by text searches constrained to subtypes of one or more of the refineable characteristics.

#### Example:

One of the defining characteristics the Concept | total replacement of hip | is | using | = | hip prosthesis. | The Concept | total replacement of hip | could be refined by allowing the user to specify one of the subtypes of "hip prosthesis."

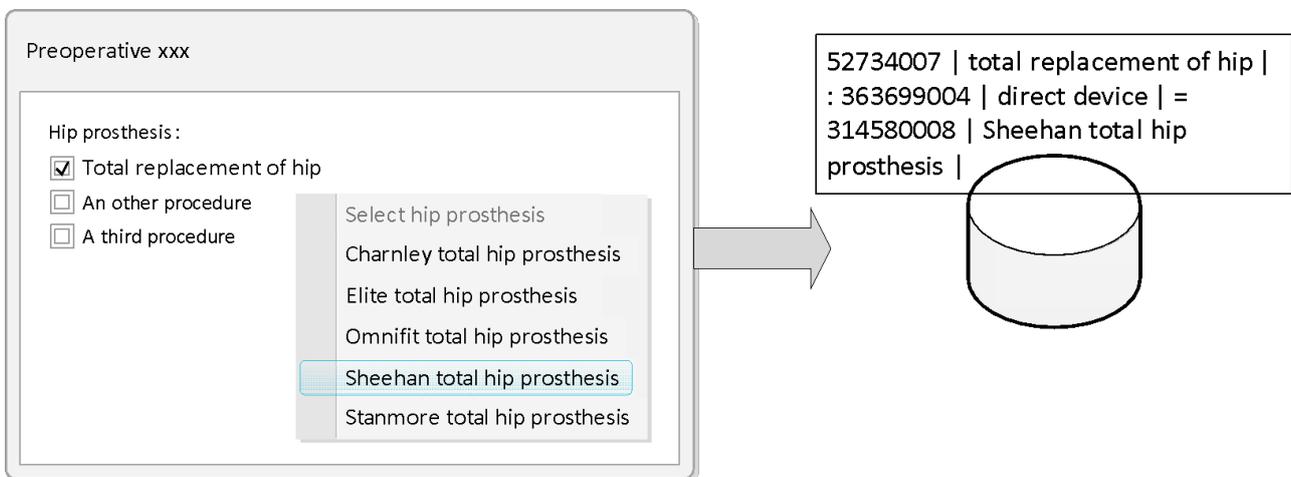


Figure 6.4.1-1: Entry of refined defining characteristics

## 6.4.2. Refinement of Concept Model Attributes

The application should also allow a user to refine the meaning of a selected Concept by selecting and applying a value to one of the attributes applicable to that type of concept. The set of permitted attributes is defined by the SNOMED CT concept model. The concept model also defines the range of values that can be applied to sanctioned attributes

### Example:

For a clinical finding such as | otitis media | it is possible to specify different values, for the following attributes | severities |, | episodicities | and | courses |.

The illustration below show how a user interface can support entry of qualifier values, and how the selection can be represented by a single SNOMED CT expression.

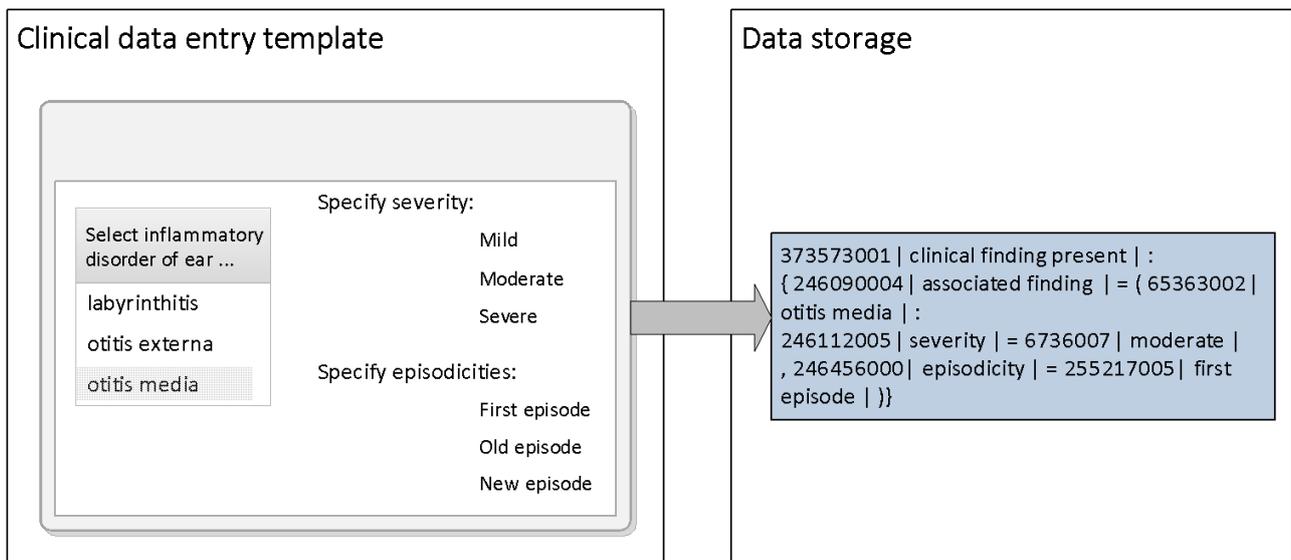


Figure 6.4.2-1: Refinement using a of data entry template

## Constraints on the Entry of Invalid Qualifiers

Qualifiers must only be used where the result of applying them results in a true subtype of the original Concept. Therefore, qualifiers should **not** be used for the purposes listed in Table 6.4.2-1. These and similar major modifications need to be handled in ways that are explicit and ensure that queries and decision support protocols are able to accurately retrieve and analyze the available information.

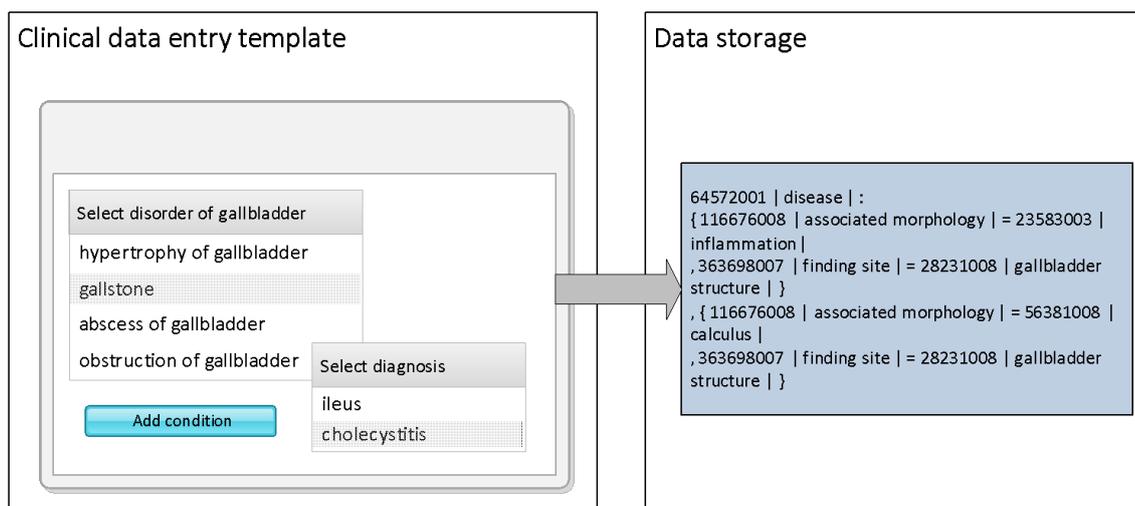
**Table 6.4.2-1: Deprecated Uses of Qualifiers**

Deprecated Use	Example	Notes
<b>Negation</b>	Fracture of humerus   must not be qualified by "excluded."	It would be inappropriate for data retrieval to treat this as a subtype of the clinical finding   Fracture of humerus  .
<b>Certainty</b>	Carcinoma of cervix   must not be qualified by "possible."	It would be inappropriate for data retrieval to treat this as a subtype of the diagnosis of   Carcinoma of cervix  .
<b>Subject of information</b>	Diabetes mellitus   should not be qualified by "family history."	It would be inappropriate for data retrieval to treat this as a subtype of the diagnosis of   Diabetes mellitus   in the patient.
<b>Planning stage</b>	Hip replacement   should not be qualified by "planned" or "requested."	It would be incorrect for a count of "Hip replacement" operations performed to include this. Decision support protocols should not assume the patient has had this operation.

### 6.4.3. Entry of Concept Combinations

Combinations of Concepts in a single expression for data entry may be required in cases where a Concept that represents the full scope of an activity is not available. Facilities for combining Concepts in an expression should be implemented and used with care.

It is appropriate to use these facilities when the combined result is conceived as a single statement that could potentially be used in many different patient records. For example, a diagnosis of "gallstones with cholecystitis" could be entered by selecting the 235919008 | gallstone (disorder) | and then selecting 76581006 | cholecystitis (disorder) | and combining these in a single statement .



**Figure 6.4.3-1: Entry of concept combinations**

It is not appropriate to use these constructs to attempt to express an entire encounter, episode or clinical history in a single statement. For example, if a patient is treated for "gallstones with cholecystitis" diagnosed by "ultrasonography of biliary tract" with a course of | amoxicillin | followed after the acute phase has resolved by a | cholecystectomy |, this should **not** be entered as a single complex post-coordinated statement combining the diagnosis, investigation and treatments.

---

 There is also a pre-coordinated Concept | Calculus of gallbladder with cholecystitis | which is equivalent to this post-coordinated combination.

## Appendix

### Appendix A - Using Search Index Files in the SNOMED CT Toolkit

This appendix focuses on searches using two indexes (or tables), which are included in the SNOMED CT Developer Toolkit. [see [SNOMED CT Developer Toolkit Guide](#)] These indexes are designed to facilitate development of effective search facilities while reducing duplication of effort. Although indexes are available in the Toolkit, implementers may need to know how to add keyword entries for any locally generated descriptions added as part of an Extension. Please refer to the "Generating the single key index" and "Generating the DualKey index" sub-sections in the Technical Implementation Guide for technical guidance on generating these indexes.

#### Search Using Single Keyword Index

The DescWordKey table is a single keyword index which is intended to enable searches which are independent of the order in which words appear in a Description. It represents the minimum necessary supporting structure for searches on SNOMED CT content. The single keyword index provides a pointer from each keyword used in any Description, to the Descriptions in which that keyword is used. This table is only useful for applications that utilise the English international release only. Note that some words that are used in *Description* are stop words, which are unlikely to be in the target of a search. These words are not considered to be *keywords* and may be excluded from the *keyword* index.

A single *keyword* search may be conducted as follows:

- The user-typed search string is converted to consistent case;
- The string is parsed, breaking at spaces and punctuation characters;
- One word is selected from the parsed word list to use as a look-up on the single keyword index;
- Look-up on the single keyword index may be "exact" or "starts with," depending on wild card conventions used in the search string.

#### Example: Search using single key-word index

The user searches for "Hip\* replacement\*" (where "\*" represents the wild card for any number of extra characters).

- The user-typed search *string* is converted to consistent case.

"Hip\* replacement" ? "HIP\* REPLACEMENT\*"

- The *string* is parsed, breaking at spaces and punctuation characters.

1. "HIP\* REPLACEMENT\*" ? (1) "HIP\*"

2. (2) "REPLACEMENT\*"

- Look up "HIP" on the single *keyword* index using "starts with" *query*.

**Table 1: Example results for a Search for "hip"**

Count	Description Identifier	Concept Identifier	Term
1	49926016	29836001	hip
2	196344018	24136001	hip
3	2296013	736004	abscess of hip
4	1480791012	386649003	partial hip replacement by prosthesis
.....	.....	.....	.....
315	371616001	1210239015	methenamine hippurate 1g tablet

Descriptions in the search results are converted to consistent case and screened, to see if they contain any words starting with "REPLACEMENT" - only those terms that do are included in the final search results. Using a Dual Key index is more efficient as the same search finds only 11 matches.

Searches involving target words that appear in many Descriptions may be unacceptably slow if searches are carried out using the single keyword index alone. Developers wishing to produce applications with faster search times are encouraged to supplement their system with a multiple keyword index.

## Search Using Multiple Keyword Index

The performance of single *keyword* searches is highly dependent on the number of candidates. *Descriptions* returned by the *keyword* for subsequent filtering. The extremely high number of matches for some words in common use makes it likely that some searches will be unacceptably slow.

One way to alleviate this problem would be to create a table containing rows for all combinations of word pairs in each *Description*. In some database environments that support optimization of multiple key searches, this may offer no benefits. However, in other environments, such a table may substantially speed searches. A comprehensive word pair table would be very large. Such a table covering the full content of *SNOMED CT* would contain approximately 1.5 million unique word pairs and 6 million rows.

Utilizing a multiple keyword index as a multiple keyword index limits the unique keys to the first three letter of each word thereby reducing the table size of the index to a more readily optimized set of keys. This requires the final part of the search to be conducted using text comparison (since the keys are incomplete).

A search on the dual key index can only be carried out if the user enters a search *string* that contains at least two word fragments both of which are three characters or more in length. If the search *string* does not meet this criterion, the single *keyword* search mechanism must be used.

- The user-typed search *string* is converted to consistent case;
- The *string* is parsed, breaking at spaces and punctuation characters;
- For each word of three characters or more, extract the first 3 characters, and arrange the word fragments in alphabetical *order*;
- Create a dual key by concatenating the first two 3 letter word fragments;
- Use this dual key to look up exact matches on the word pair index;
- *Descriptions* found by searching on the word pair index are screened, to see if they contain the complete words in the original search *string*

### Example: Search using word pair index

User searches for "PYRO\* 1 OXYGEN\*".

The *string* is parsed, breaking at spaces and punctuation characters.

1. "PYRO\*";
2. 1;
3. "OXYGEN\*".

For each word of three characters or more, extract the first 3 characters, and arrange the word fragments in alphabetical *order*.

1. "OXY";
2. "PYR".

Create a dual key by concatenating the first two 3 letter word fragments.

OXYPYR

Use this dual key to look up exact matches on the word pair index.

**Sample results of a search for "PYRO 1 OXYGEN\*\*"**

Dual key	Description Identifier	Description
OXYPYR	1969019	2,5-Dihydroxy-pyridine oxygenase
OXYPYR	22565018	pyrogallol 1,2-oxygenase
OXYPYR	104951019	2,5-Dihydroxy-pyridine oxygenase

Descriptions found by searching on the word pair index are screened, to see if they contain the complete words in the original search string:

- Description, 1969019, is eliminated since it does not contain the word "1";
- Description, 104951019, is eliminated; it does not contain the word "1" or any word beginning with the string "pyro".

## Appendix B - Future Additions to this Guide

A set of techniques to support effective and efficient SNOMED CT search and data entry have been described in the Guide. However, more techniques could be included in future. The following lists of techniques are being considered in future releases of the SNOMED CT Search and Data Entry Guide.

1. Searching and displaying:
  - a. Superscript (^ ^) and subscript (> <) (e.g. for chemical formulae or radionuclides <sup>186</sup>Platinum)
  - b. Diacritic and accented characters (e.g. Sjögren can be found by entering 'sjogren')
  - c. Hyphens (e.g. "dot-blot hemorrhage of superior aspect of far peripheral retina of left eye" can be found using by entering "dot blot")
2. Case sensitive matching
3. Using wild cards
4. Search functionality identifying discrete tokens in both user search expressions and in descriptions being indexed as those substrings separated by any of: <space><tab>()[]/.,:;%#&+~'^><="
5. Number substitutions (e.g. "2" can be found when entering "two" or II, or vice versa).
6. An enhancement to "Search for words within a description in any order" sub-section: Stripping off last word of the search term
7. Constraining searches to at least 3 characters in a search term
8. Extending search by stemming substitution
9. Show search results as they are received not wait until they are completely received

## Appendix C - An Example Search and Data Entry Overview Table

This section presents an example of a table, which is currently under development, intended to provide an overview of the various search techniques presented in this document, and their applications.

### Search by Term Text – Overview Table

Table assumptions:

- Search terms are at least 3 characters (what about phrases like 'of' or very short terms like "mg" or "MI")
- Filters should be applied
- Case sensitivity
- Search term preparation (and maybe SNOMED CT term preparation)

Ordering and presentation have not been considered.

	<b>Example search scenario</b>	<b>Environment</b> ? - Very useful ? - not very useful		<b>Potential Number of results</b>	<b>Implementation Complexity</b>
		Browser	Clinical		
Exact phrase	When the user knows the exact term string that he requires	?	?	Low	Simple

Search for descriptions that begin with the search text	User needs to find terms where they know the root of the word/words they are looking for but not the order that it might be in in a term string.	?	?	Medium	Moderate
Search for descriptions that contain the search text	When the user somewhat knows the term string that they require but not the full description	?	?	High	Simple
Search for descriptions that end with the search text	When the user may want to search for the required text at the end of words	?	?	Low	Moderate
Search for words within a description in any order	When the user does not how words are ordered in the Descriptions.	?	?	High	Moderate
Search for identical terms	When the user knows the exact term string that they require	?	?	Low	Simple
Search for word(s) in a specific order (or a matching phrase)	When the user somewhat knows the term string that they require but not the full description	?	?	Medium	Moderate
Search for Concept Identifiers	When the user wants to know the Description of the Concept that he requires	?	?	Low	Simple
Search for Description Identifiers	When the user wants to knows the Concept of the Description that he requires	?	?	Low	Simple