Apelon Distributed Terminology System (DTS)

DTS Matching – Quick Overview

Table of Contents

Introduction

The API for Apelon DTS provides the capability to search **namespaces** for concepts and terms that include the search text strings that you specify (a namespace refers to a specific vocabulary source in your knowledgebase). You can filter search results so that only concepts that meet your search criteria are returned (these criteria include concepts with selected roles, properties, and associations).

The API also includes the capability to **match** search strings to concept terminology within a specific data repository from your knowledgebase. Each of these repositories is referred to as a "**silo**."

Selector & Extractor

The **Selector & Extractor** utility, included with Apelon DTS, creates these silos and populates them with terms and associated concepts from one or more namespaces in your knowledgebases. Using silos allows you to optimize searching capabilities when you use Apelon DTS to perform searches of your knowledgebase.

The concept data acquired from the knowledgebase to populate each silo is based on one or more sets of customized filters. The **Selector & Extractor** uses specifications from the **filterSpec.xml** configuration files to select and extract data from the knowledgebase to populate the silo(s). You can customize the data in the silos by modifying the installed **filterSpec.xml** file, then running the Selector & Extractor to populate the silos with the new data. You also can create additional **filterSpec.xml** files (each with a unique name). Each time you run the Selector & Extractor, the data in existing silos will be updated, and new silos will be created, based on the parameters in the **filterSpec.xml** files.
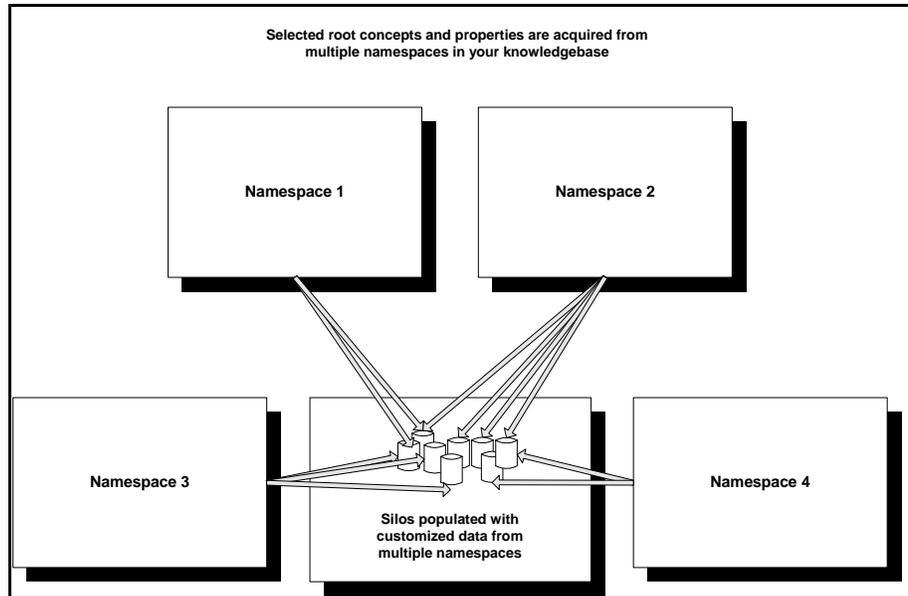
MatchPack

You can customize the precision by which matching will occur between your search string and the concepts in the silo you choose for searching. You control this matching precision based on a series of increasingly tolerant matching options, called the **MatchPack** matching options.

Only concepts, or concept synonyms, with names that meet the matching criteria are retrieved from the search of the silo. The number of concepts retrieved from each search depends on which combination of matching options you select.

The DTS Browser Web client includes the MatchPack matching capabilities. You can perform **Advanced** and **Quick** searches within a specific silo based on your customized MatchPack option selections. The DTS Editor also provides search options that include the MatchPack matching capabilities.

Searching For Matches in a Data "Silo"

Apelon DTS allows you to search a specific knowledgebase **silo** for concepts and related terms that match the search strings you specify. A single silo might be defined, or there may be multiple silos. Each silo is a repository of customized, semantically related concepts (e.g., all descendants of the concept **Disease**) and extracted character strings associated with these concepts. This data can be from multiple namespaces within your knowledgebase.

The Apelon DTS servers attempt to match the search strings that you submit against the clinical terms and associated concepts in a specific silo. Each unique concept in a silo is associated with a set of alternative terms. A search string must match a term in the knowledgebase (or its lexical variant[s]) in order to be matched by the server.

- The concept data acquired from the knowledgebase to populate each silo is based on the parameters in one or more configuration files.

- The **Selector & Extractor** utility generates silos and populates them with data from the knowledgebase.

- You run the **Selector & Extractor** utility by executing **selext.bat** on the Command Line (the **selext.bat** file is located in the installed *DTSInstall***bin\selext** directory, where *DTSInstall* represents the current DTS version).

> To run the **Selector & Extractor** utility for a Linux installation, execute **selext.sh** on the Command Line; the **selext.sh** script is located in the installed **bin/selext** directory)

  - The **Selector & Extractor** uses specifications from the **filterSpec.xml** file to select and extract knowledgebase data to populate the silo(s).

  - You can modify the installed **filterSpec.xml** file to reflect data you want to acquire from your knowledgebase, or create additional **filterspec.xml** files (each with a unique name).

  - You then must run the Selector & Extractor to generate and populate the customized silos.

    - ◊ On the Command Line, execute the following:
      selext/filterspec.xml
      selext/filterspec2.xml
      selext/filterspec3.xml

    - ◊ Run the Selector & Extractor *only* after you have completed modifications to all **filterSpec.xml** file specifications.

- Refer to the *Knowledgebase Administrators Guide* for more on the Selector & Extractor utility, and for detailed instructions on customizing the filterSpec.xml files.

MatchPack Matching Options

The API for Apelon DTS allows you to match search strings to knowledgebase concept terminology based on a series of matching options, referred to as **MatchPack** matching options. Matches are attempted in the order indicated in the following discussions.

Based on the option(s) you choose, silo searches retrieve results that are "ranked." The results with the fewest matched **tokens** are listed first (i.e., concepts that satisfy the matching criteria and containing the **fewest extraneous words** are returned first). Note that the use of wildcards (*****) is not supported for silo searches.

The DTS Browser includes these matching capabilities to allow you to perform **Advanced** and **Quick** searches within a specific silo. When matched concepts are returned from the search, you can navigate to any related concept in the knowledgebase.

Complete Match

This is the most exact method of matching the text search string that you specify to either a concept name, or the name of one of its synonyms, in the silo you selected. Complete Match is the default matching method, and is always active (even if you choose one of the other, more tolerant, matching methods).

- Through Complete Match, a match occurs and a concept is retrieved when either the concept name, or the name of one of its synonyms, in the silo contains the same words as the search string you specify, in any order.

- Every concept in the silo with a name, or a synonym with a name, that contains all the words in the search string (in any order) is retrieved.

Under Match

If Under Match is activated, a match occurs and a concept is retrieved if the concept name, or the name of one of its synonyms, in the silo matches the search string words (in any order) *and the concept or synonym name contains words in addition to those in the search string*.

- When Under Match is selected, the order of the words in the search string is not considered in the search (e.g., the identical results are returned for the search strings **myocardial infarction** and **infarction myocardial**).

- You can use the **Best Match Only** option in conjunction with Under Match.

    o If **Best Match Only** also is active, Complete Match is performed first (i.e., if the concept name, or the name of one of its synonyms, in the specified silo contains every word in the search string, in any order, a match occurs).

    o If one or more complete matches are found, *only* these (complete) matches are returned, and no further matching attempt is made.

- o If no concept matches occur from the Complete Match, Under Match is attempted; if under matches are found, *only* these matches are returned.

- o If Under Match is activated, and **Best Match Only** is inactive, under matches will be returned *in addition to* the complete matches.

Partial Under Match

If Partial Under Match is activated, a match occurs and a concept is retrieved if the concept name, or the name of one of its synonyms, in the specified silo contains **all remaining words** (in any order) after all **stopwords** have been removed.

- • The following stopwords are those words that are considered as unnecessary for the search:

  A
  and
  by
  for
  in
  NOS
  of
  on
  the
  to
  with

- • Example: For the search string **attack in the heart and brain**, the string becomes **attack heart brain** because the stopwords **in**, **the**, and **and** are discarded. If the words in the modified search string match a concept name, or the name of one of its synonyms, in the silo (in any order) a match is returned.

- • You can use the **Best Match Only** option in conjunction with Partial Under Match.

  - o If **Best Match Only** also is active, Complete Match is performed first (i.e., if the concept name, or the name of one of its synonyms, in the specified silo contains every word in the search string, in any order, a match occurs).

  - o If one or more complete matches are found, *only* these (complete) matches are returned and no further matching attempt is made.

  - o If no concept matches occur from the Complete Match, Under Match is attempted; if under matches are found, *only* these matches are returned.

- o If no matches occur from Under Match, matches are attempted through Partial Under Match.
- o If Partial Under Match is activated, and **Best Match Only** is inactive, partial under matches will be returned *in addition to* the complete matches and under matches.

Spell Checking

Spell Checking is one of the MatchPack options you can use to search a specified silo. In the event a misspelled search string is entered, spell checking generates suggested spelling variants against which matches can be attempted. Spelling correction will occur when MatchPack attempts to match a concept name, or the name of a concept synonym, with the search string you specify.

- You can use spell checking regardless of the matching option you select (**Complete Match**, **Under Match**, or **Partial Under Match**).
- The number of matched concepts retrieved from the search is dependent on the matching method you select.

[Back to Top](#)