CAS
A DIVISION OF THE
AMERICAN CHEMICAL SOCIETY

ClientServices

Phone: 800-631-1884, 614-447-3870
Fax: 614-447-3747
E-mail: answers@cas.org
Web: www.cas.org/products/other-cas-products/client-services

Chemical Name Match

Input Data

The name match algorithm determines from context which characters should be uppercase or lowercase, italic, subscript, superscript, or even small capital. Therefore, it is not necessary to input characters in any way other than as noted above, using either uppercase or lowercase alphabetics.

Optimizing Name Match

To get the best results out of Name Match processing, it is important to take into account the nature of the names in CAS REGISTRY^{SM}, and the programs that perform the matching of the names.

Each substance in CAS REGISTRY is associated with a CA Index Name. These names are prepared by CAS experts using very specific rules to assure the assignment of one unique name for each substance. This is usually a highly systematic, chemically descriptive name.

In addition to the CA Index Name, many substances also have other names (synonyms) associated with them. Some of these are systematic and chemically descriptive. Others are less descriptive or trivial or trade names.

Specific examples of the types of names available for name match are listed in the four categories below. Since those in the first category (A) are the most chemically specific, they might appear to be the most reliable for name matching. However, due to the large number of possible variations, they are not. The types of names in the other three categories will usually provide better name match results.

Types of Names in the CAS Registry System

A. Systematic names that are long, complex, or have a relatively high degree of punctuation.
    a. Examples
        i. 2-naphthalenecarboxylic acid, 3-hydroxy-4-[(4-methyl-2-sulfophenyl)azo]-, calcium salt(1,1)
        ii. 1-Phenyl-2,3-dimethyl-5-pyrazolone
B. Systematic names that have a low degree of punctuation, or natural product names with minimal substitution.
    a. Examples
        i. p-Aminobenzoic acid
        ii. D-Lactic acid

C. Names that contain no punctuation. These include many esters, salts, natural products, and names used in trade.
    a. Examples
        i. Quinidine sulfate
        ii. Sodium salicylate
        iii. Parathion
D. Names that end in numbers or contain two or more capital letters. Most of these names are acronyms, dyes, trade names, lab numbers, or natural product names.
    a. Examples
        i. EDTA
        ii. Penicillin G
        iii. BA 2685
        iv. Direct Blue 3B

Specific Suggestions for Optimizing Name Match Results

- **Invert Index Names.** Systematic names appear in the Registry File in either inverted or uninverted form. CA Index Names, however, are always stored in the inverted form. If a candidate name is known to be a CA Index Name, better results will be obtained if it is inverted.
    - *2-Hydroxy-1,4-naphthalenedione* will have the inverted Index Name *1,4-Naphthalenedione, 2-hydroxy-*
- **Drop the word "of."** Better results will be obtained if names that include phrases such as "ether of," or "hydrochloride of," or "oil of" are reformatted to omit the word "of:"
    - *Wintergreen oil* **not** *Oil of wintergreen*
    - *Diethylamine hydrochloride* **not** *Hydrochloride of diethylamine*
- **"Americanize" names.** Although there are many spelling variants in the Registry File, the following rules will generally improve retrieval results.
    - Names from foreign languages, following transliteration, should be altered to "Americanize" the spellings.
        - *Hemoglobin* **not** *Haemoglobin*
        - *Glycol* **not** Glykol
        - *Sulfur* **not** *sulphur*
    - Alkaloid names ending with "in" should be changed to "ine"
    - Glycoside names ending with "ine" should be changed to "in"
    - Steroid names ending with "an" should be changed to "ane" if no unsaturation is present. If saturation is present, "an" should be kept.
- **Simplify alphanumeric and laboratory designations.** These types of names will be more likely to match if they conform to the following guidelines.
    - A class name should be cited ahead of the number and/or letter designation.
        - *Antibiotic FR 1923*
        - *Alkaloid AD-V*
        - *Amine 220*
    - Terms such as "compound" or "substance" are not generally used with alphanumeric or other code designations.
        - *67/20* **not** *Compound 67/20*
        - *K525* **not** *Substance K525*
    - Company names are not cited with alphanumeric or other code designations
        - *GC 4072* **not** *General Chemical 4072*

        - *BAY 5097* **not** *Bayer 5097*
    - Process and use information should not be associated with names. Descriptive terms such as grade, source, and physical state should not be included.
        - *Silica* **not** *Pyrogenic silica*
        - *Limestone* **not** *Ground limestone*
        - *Sulfur* **not** *Sulfur (amorphous)*