NIST Text Format of Individual Spectra

The NIST Text file format is simple. A file can contain as many spectra as wanted. Each spectrum must start with the Field title "Name:". There must be something in this field in order for the spectrum to be imported (i.e., the data file name and the scan number of the spectrum's source). The next required Field title is "Num Peaks:". The contents of this field must be the number of mass/intensity pairs that make up the spectrum. Optional fields with the titles of "Comments:", "Formula:", "MW:", "Synonym:", and "CAS#:" can be between the "Name:" and "Num Peaks:" fields.

When a spectrum is saved to a text file from the **NIST/EPA/NIH Main** or **Replicates Database**, it will also contain fields with the names "NIST#:" and "DB#:". The field with the name "NIST#:" is on the same line as the "CAS#:" field and separated by a semicolon (;). "DB#" is same as the "ID#" displayed in the **Spectrum Text** Window. The mass/intensity list begins on the line following the line with the Field title "Num Peaks:". The lines of the mass/intensity list have no Field title.

When a file is created with all the allowable fields, each Field title should be on a separate line. Use the following format. The format for each spectrum in a text file usable by the **Program** should be as follows:

Line 1: **NAME: Molecule** (Required, up to 511 characters)
Line 2: **COMMENT: Run 23, 8/8/88** (Optional, up to 1023 characters)
Line 3: **FORMULA: C6H6** (Optional, up to 23 characters)
Line 4: **MW: 78** (Optional)

If the **User spectrum Search** contains a CAS registry number with its appropriate prefix (**CAS:**), the display of the spectrum in the **Plot**, **Compare**, and **Structure** Windows will have the structure of the compound with that CAS registry number if it is present in the **NIST/EPA/NIH Main Database**.

Line 5: **CAS: 71-43-2** (Optional)
Line 6: **SYNONYM: Chemical name synonym** (Optional, may be repeated)

The actual mass spectral data (number of peaks in the spectrum) must begin on the next line. It does not actually matter what line this is, as long as it precedes the line that starts the mass/intensity pair data. The VERY NEXT line and subsequent lines MUST contain the paired mass/intensity values.

Line 7: **Num Peaks: 18** (This prefix and the exact number of mass/intensity pairs is Required.)
Line 8: **26 430; 27 340; 28 40; 37 480; 38 611; 39 1411; 49 300; 50 1792;**
Line 9: **51 2052; 52 1962; 63 340; 73 160; 74 480; 75 180;**

…

Line n: **76 721; 77 1401; 78 9806; 79 651;**

The peaks need not be normalized, and the masses need not be ordered. The exact spacing and delimiters used for the mass/intensity pairs are unimportant. For example, the peaks above could equally be presented as:

(26,430),(27,340),(28,40),(37,480),(38,611),(39,1411),
(49,300),(75,180),(50,1792),(51,2052),(52,1962),(63,340),
(73,160),(74,480),(75,180),(76,721),(77,1401),(78,9806),(79,651)

or each mass intensity pair could be on an individual line. The following characters are accepted as delimiters (except "|"): `|space|tab|,|;|:|(|)|[|]|}|`

The file can be given any valid file name; however, it is best to use the extension "MSP". This is the recognized default extension in the **MS Search Program** when user-database spectra are to be imported.

Several examples of NIST text format mass spectral files are installed together with the **MS Search Program**, namely SAMPLIB.MSP, UNKNOWN.MSP, and USERDEMO.MSP.