Liberator

Table of contents

1. Overview

This application creates a library of annotated spectra generated from a pepxml file.

2. Usage

```
usage: Liberator <PepXML> [-a] [--auto-scan] [-b <arg>] [-d <arg>] [-e
       <arg>] [-f <arg>] [--filter <arg>] [--force] [-g <arg>] [-h] [-i
<arg>] [-l]
       [--loss-advanced <arg>] [--loss-phosphate] [--loss-verbose]
       [--max-loss-sites <arg>] [--max-peak-annots <arg>] [--maxCharge
<arg>]
       [--maxMZ <arg>] [--mgf-scan-pattern <arg>] [--minCharge <arg>]
[--minMZ
       <arg>] [--msLevel <arg>] [-n] [--no-merge] [-o <arg>] [-p <arg>]
[-q] [-r <arg>] [--tol]
       [-v]
 -a,--average                  set the average mass mode for peptide mass
                               calculation
                               by default: monoisotopic.
    --auto-scan                enable auto scan number (from scan 1 to n).
 -b,--bSize <arg>              batch size for MS files
                               by default: 10000.
 -d,--dir <arg>               define the resource directory(ies) ('path1,
..., pathn') from where
                               MS data files will be recursively searched.
The basename file is defined in
                               each "spectrum" attribute of
'<spectrum_query>' tag with the following format
                               ('basename.start_scan.end_scan.charge')
                               by default: empty path (a unique MS resource
file path is defined for each
                               '<msms_run_summary>' as defined in
http://sashimi.sourceforge.net/schema_revision/pepXML/pepXML_vXYZ.xsd).
 -e,--ext <arg>                extension for MS2 spectrum files
                               by default: mgf.
 -f,--frags <arg>              define the fragmentation type(s):
                               a,b,c,x,y,z for backbone type fragments
                               i for immonium and p for precursor
                               by default: b,y.
    --filter <arg>             set the filter for peptide to spectrum
search hit selection.
                               The expression operands are defined from
search_score 'name' and
                               analysis_result 'analysis' pepxml attribute
value (ex:
                               (peptideprophet # 0.9) & (hyperscore > 400))
                               by default: peptideprophet # 0.9.
    --force                    force the conversion of already converted
                               MS files (MS data files serialized in /tmp).
```

```
 -g,--graph <arg>               export PSM graph file.

 -h,--help                      print this message.
 -i,--setting-file <arg>        give a property file with all input
                                settings.
 -l,--loss                      enable neutral loss fragmentation.

    --loss-advanced <arg>       enable the user to define losses himself.
                                Set the amino-acid target and the kind of
loss action based on the grammar
                                below:
                                // Examples of Rules
                                # simple:
                                C -> -16
                                # AA class target:
                                [KNQR] -> -23.5
                                # AA class with a modif range as target:
                                [KNQR]/40:80 -> del
                                # alternative editions:
                                [KNQR]/40:80 -> -23.5, del
                                # few rules:
                                [KNQR]/40:80 -> -23.5, del ; C -> -16
                                # sequential edit actions:
                                Y/79.9:80.1 -> del-18 ;
                                // Grammar Definition
                                rules:         rule (';' rule )*;
                                rule:          target '->'  actions;
                                target:        aaclass ('/' range) ?;
                                actions:       action (',' action)*;
                                action:        (modaction)+ | remaction
(modaction)*;
                                modaction:     '-'
                                INT | '-' FLOAT;
                                remaction:     'del';
                                aaclass:       AA | '['  AA (AA)* ']';
                                range:         number  ':' number;
                                number:        INT | FLOAT;
                                // Lexers
                                INT:           '0'..'9'+ ;
                                FLOAT:         INT ('.' INT);
                                AA:            'A'..'Z' ;
    --loss-phosphate            enable phosphate loss on phosphated [YST]
                                amino acid fragments.
                                by default: false.
    --loss-verbose              enable loss rule verbosity mode.
                                by default: false.
    --max-loss-sites <arg>      limit the maximum of loss sites
                                by default: 2
    --max-peak-annots <arg>     limit the maximum of annotations displayed
                                by peak
                                by default: no limit
    --maxCharge <arg>           maximal precursor charge
                                by default: 10.
    --maxMZ <arg>               maximal precursor mz value
```

```
                                by default: 5000.0.
   --mgf-scan-pattern <arg>     set the pattern to extract scan number from
                                mgf TITLE tag
   --minCharge <arg>            minimal precursor charge
                                by default: 1.
   --minMZ <arg>                minimal precursor mz value
                                by default: 100.0.
   --msLevel <arg>              level of ms spectra
                                by default: 2.
-n,--no-clobber                 do not overwrite existing converted MS
                                files (MS data files serialized in /tmp).
   --no-merge                   disable the merging of MS spectra matching
                                the same peptide/z.
-o,--output <arg>               set the output filename

-p,--precision <arg>            define the number of fractional digits for
                                output
                                by default: 4.
-q,--quiet                      quiet mode (verbose off)
                                by default: false.
-r,--priority-rule <arg>        set the annot precedence rule for
                                multi-annotated peaks:
                                // Grammar Definition
                                rules:         rule ('>>' rule )*;
                                rule:          rule-ion | rule-num
                                rule-ion:      'ion' : type-seq
                                rule-num:
'loss'|'charge'|'mzdiff':(ascent|descent)?
                                type-seq:
                                ion-type ('>' ion-type)+
                                ion-type:      'a' | 'b' | 'c' | 'x' | 'y' |
'z'

                                by default: charge>>ion:y>b>>loss>>mzdiff.
   --tol                        define the tolerance for mz fragment peak
                                comparison
                                by default: 0.1.
-v,--version                    print the version info.
```

3. Releases

The latest version v1.47 - Download app with the default properties file.

3.1. Rel1.47

Type	Summary
Bug fixed	Serialized MS data was corrupted when scan number was exceeded 99999 (scan number limit is now of 1016-1).

3.2. Rel1.46

Type	Summary
New	Param *--no-clobber* prevent the default overwrite of existing converted MS files.
New	Param *--no-merge* prevent the merge of MS spectra that match the same peptide/z.

3.3. Rel1.44

Type	Summary
Change	Liberator interrupted the library spectra process while building an ambiguous peptide. Such a peptide contains ambiguous or undefined amino-acid(s), preventing the computation of correct fragmentation spectrum. It now just warns the user.

3.4. Rel1.43

Type	Summary
Change	Param *--dir* now supports many directory resources like in "dir1, dir2".
Bug fixed	In release 1.42, annotated spectra were hugely duplicated in the created library.

3.5. Rel1.42

Type	Summary
New	New parameter *--force* to force the reconversion of MS data files.
Bug fixed	MGF reader bug: a spectrum builder exception was thrown while reading some MGF files with multiple retention times.

3.6. Rel1.41

Type	Summary
New	New parameter *--auto-scan* auto increment scan numbers for MS data.
New	New parameter *--mgf-scan-pattern <arg>* set the regular expression to extract scan number from TITLE tag (mgf format only).

3.7. Rel1.4

Type	Summary
New	New parameter *--max-peak-annots <arg>* limiting the maximum number of annotations by peak.
Change	Change the grammar of parameter *--priority-rule*.

3.8. Rel1.3

Type	Summary
Bug fixed	Some problematic MGF entries with overlaps between "SCANS" leading to crash while reading the pepxml.

3.9. Rel1.2

Type	Summary
New	New parameter *--max-loss-sites <arg>* limiting the maximum number of simultaneous loss by fragment.
New	Enable progress bar for MSn processing.
New	New parameter *--loss-verbose* enabling verbose mode for loss rule specifications.

3.10. Rel1.1

Type	Summary
New	New parameter *--loss-phosphate*.

3.11. Rel1.0

Type	Summary
New	New parameter *--loss-advanced*.

3.12. Rel0.6

Type	Summary
Runtime	Task name in progress bar (left margin)
Bug fixed	mixed-mode: merge msms_run_summary(s) in one (no more peptide redundancies)

3.13. Rel0.5

Type	Summary
Runtime	Support multi-resources MS data. The file basename is coded in the attribute 'spectrum' of the pepxml element 'spectrum_query'.

3.14. Rel0.2

Type	Summary
Change	Change the --pp-threshold <numeric> with --filter <inequation expression>.