Package 'IDSL.MXP'

March 24, 2023

Type Package

Title Parser for mzML, mzXML, and netCDF Files (Mass Spectrometry Data)

Version 2.0

Depends R (>= 4.0)

Imports xml2, base64enc

Suggests RNetCDF

Author Sadjad Fakouri-Baygi [aut] (<https://orcid.org/0000-0002-6864-6911>), Dinesh Barupal [cre, aut] (<https://orcid.org/0000-0002-9954-8628>)

Maintainer Dinesh Barupal <dinesh.barupal@mssm.edu>

Description A tiny parser to extract mass spectra data and metadata table of mass spectrometry acquisition properties from mzML, mzXML and netCDF files introduced in <doi:10.1021/acs.jproteome.2c00120>.

License MIT + file LICENSE

URL https://github.com/idslme/idsl.mxp

https://colab.research.google.com/drive/1gXwwuI1zzDHykKfodLSQQt5rwTuFEMpD

BugReports https://github.com/idslme/idsl.mxp/issues

Encoding UTF-8

Archs i386, x64

NeedsCompilation no

Repository CRAN

Date/Publication 2023-03-24 10:00:16 UTC

R topics documented:

`getNetCDF` *getNetCDF*

Description

This function returns a list of two data objects needed for the mass spectrometry data processing.

Usage

```
getNetCDF(MSfile)
```

Arguments

`MSfile`	name of the mass spectrometry file with .cdf extension

Value

`scanTable`	a dataframe of different scan properties including 'seqNum', 'msLevel', 'polarity', 'peaksCount', 'totIonCurrent', 'retentionTime', 'basePeakMZ', 'basePeakIntensity', 'collisionEnergy', 'lowMZ', 'highMZ', 'precursorScanNum', 'precursorMZ', 'precursorCharge', 'precursorIntensity', 'injectionTime', 'filterString', 'scanType', 'centroided', 'isolationWindowTargetMZ', 'isolationWindowLowerOffset', 'isolationWindowUpperOffset', 'scanWindowLowerLimit', and 'scanWindowUpperLimit'.
`spectraList`	a list of matrices of m/z and intensity values for each chromatogram scan

Note

'retentionTime' column in the 'scanTable' object is presented in minute.

`getScanTable` *getScanTable*

Description

This function creates a scanTable from chromatogram scans of the mass spectrometry data.

Usage

```
getScanTable(xmlData, msFormat)
```

Arguments

`xmlData`	A structured data of the mass spectrometry data created by the 'read_xml' function.
`msFormat`	format extension of the mass spectrometry file c("mzML", "mzXML")

Value

a dataframe of different scan properties including ‘seqNum‘, ‘msLevel‘, ‘polarity‘, ‘peaksCount‘, ‘totIonCurrent‘, ‘retentionTime‘, ‘basePeakMZ‘, ‘basePeakIntensity‘, ‘collisionEnergy‘, ‘lowMZ‘, ‘highMZ‘, ‘precursorScanNum‘, ‘precursorMZ‘, ‘precursorCharge‘, ‘precursorIntensity‘, ‘injectionTime‘, ‘filterString‘, ‘scanType‘, ‘centroided‘, ‘isolationWindowTargetMZ‘, ‘isolationWindowLowerOffset‘, ‘isolationWindowUpperOffset‘, ‘scanWindowLowerLimit‘, and ‘scanWindowUpperLimit‘. ‘scanType‘ is only provided for the mzXML data format.

Note

‘retentionTime‘ column is presented in minute.

Examples

```
temp_wd <- tempdir()
temp_wd_zip <- paste0(temp_wd,"/idsl_ipa_test_files.zip")
download.file(paste0("https://github.com/idslme/IDSL.IPA/blob/main/",
"IPA_educational_files/idsl_ipa_test_files.zip?raw=true"),
destfile = temp_wd_zip, mode = "wb")
unzip(temp_wd_zip, exdir = temp_wd)
xmlData <- xml2::read_xml(paste0(path = temp_wd, "/", MSfile = "003.mzML"))
scanTable <- getScanTable(xmlData, msFormat = "mzML")
```

---

getSpectra *getSpectra*

Description

This function creates a spectraList for the chromatogram scans of the mass spectrometry data.

Usage

```
getSpectra(xmlData, msFormat)
```

Arguments

xmlData	a structured data of the mass spectrometry data created by the ‘read_xml‘ function.
msFormat	format extension of the mass spectrometry file c("mzML", "mzXML")

Value

a list of matrices of m/z and intensity values for each chromatogram scan

Examples

```
temp_wd <- tempdir()
temp_wd_zip <- paste0(temp_wd,"/idsl_ipa_test_files.zip")
download.file(paste0("https://github.com/idslme/IDSL.IPA/blob/main/",
"IPA_educational_files/idsl_ipa_test_files.zip?raw=true"),
destfile = temp_wd_zip, mode = "wb")
unzip(temp_wd_zip, exdir = temp_wd)
xmlData <- xml2::read_xml(paste0(path = temp_wd, "/", MSfile = "003.mzML"))
spectraList <- getSpectra(xmlData, msFormat = "mzML")
```

MXP_locate_regex *MXP Locate regex*

Description

Locate indices of the pattern in the string

Usage

```
MXP_locate_regex(string, pattern, ignore.case = FALSE, perl = FALSE, fixed = FALSE,
useBytes = FALSE)
```

Arguments

string	a string as character
pattern	a pattern to screen
ignore.case	ignore.case
perl	perl
fixed	fixed
useBytes	useBytes

Details

This function returns 'NULL' when no matches are detected for the pattern.

Value

A 2-column matrix of location indices. The first and second columns represent start and end positions, respectively.

Examples

```
pattern <- "Cl"
string <- "NaCl.5HCl"
Location_Cl <- MXP_locate_regex(string, pattern)
```

peak2list *Peak to List (The main function)*

Description

This function returns a list of two data objects required for the mass spectrometry data processing.

Usage

```
peak2list(path, MSfileName = "")
```

Arguments

path	address of the mass spectrometry file
MSfileName	name of the mass spectrometry file with .mzML or .mzXML extensions

Value

scanTable	a dataframe of different scan properties including 'seqNum', 'msLevel', 'polarity', 'peaksCount', 'totIonCurrent', 'retentionTime', 'basePeakMZ', 'basePeakIntensity', 'collisionEnergy', 'lowMZ', 'highMZ', 'precursorScanNum', 'precursorMZ', 'precursorCharge', 'precursorIntensity', 'injectionTime', 'filterString', 'scanType', 'centroided', 'isolationWindowTargetMZ', 'isolationWindowLowerOffset', 'isolationWindowUpperOffset', 'scanWindowLowerLimit', and 'scanWindowUpperLimit'. 'scanType' is only provided for the mzXML data format.
spectraList	a list of matrices of m/z and intensity values for each chromatogram scan

Note

'retentionTime' column in the 'scanTable' object is presented in minute.

See Also

https://colab.research.google.com/drive/1gXwwuI1zzDHykKfodLSQQt5rwTuFEMpD

Examples

```
temp_wd <- tempdir()
temp_wd_zip <- paste0(temp_wd,"/idsl_ipa_test_files.zip")
download.file(paste0("https://github.com/idslme/IDSL.IPA/blob/main/",
"IPA_educational_files/idsl_ipa_test_files.zip?raw=true"),
destfile = temp_wd_zip, mode = "wb")
unzip(temp_wd_zip, exdir = temp_wd)
p2l <- peak2list(path = temp_wd, MSfileName = "003.mzML")
```

Index