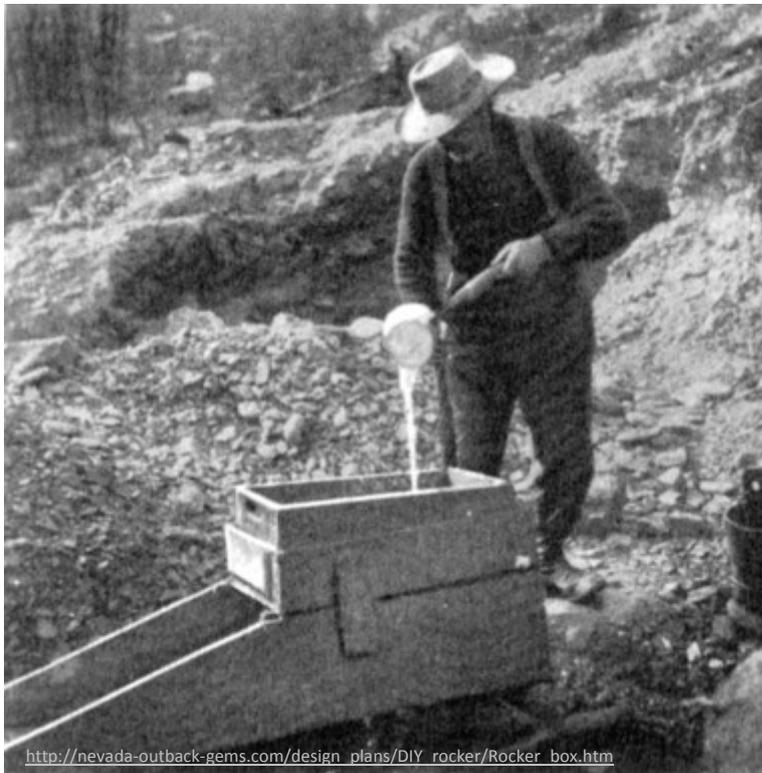


1.2.3

RockerBox user guide



Henk van den Toorn

Utrecht University, Netherlands Proteomics
Centre & Netherlands Bioinformatics Centre

10/12/2010

CONTENTS

Welcome!	3
Prerequisites	3
Hardware	3
RockerBox	3
Java	3
Percolator	3
RockerBox main window overview	4
Exporting data	5
Export Peptide List	5
Export Feature list for Percolator	5
Export Peak List (MGF)	5
Creating Charts	6
Dialog	6
Chart output	6
Filtering	7
Manual Filter	7
FDR based filter	8
Percolator based filter	9
TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION	10
1. Definitions	10
2. Grant of Copyright License	10
3. Grant of Patent License	10
4. Redistribution	11
5. Submission of Contributions	11
6. Trademarks	11
7. Disclaimer of Warranty	11
8. Limitation of Liability	11
9. Accepting Warranty or Additional Liability	11

WELCOME!

RockerBox is an application to handle large .dat files. There are several possibilities to chart your data, export data to text files or to create a smaller .dat file by filtering methods. In this document, we'll explain how to use the application. With these instructions you will be on your way to analyze large data files. So jump right in!

PREREQUISITES

HARDWARE

RockerBox has been developed and tested on fairly modest hardware (by today's standards), a machine with a dual core Intel® Xeon™ processor running at 3.0 GHz with 4 Gb RAM on 32-bit Microsoft® Windows 7™.

In order to run, RockerBox typically needs around 1.5 Gb of RAM assigned to it, so we recommend a minimum of 3 Gb memory.

ROCKERBOX

To obtain the latest version of RockerBox, go to

<https://trac.nbic.nl/rockerbox/downloads>

If you're a Windows user, choose the *RockerBox executable wrapped for Windows*. You can unpack the .exe file anywhere on your computer and double-click on it to run it.

For Apple® Macintosh™ or Linux® machines, download the *RockerBox executable*. For large files, **do not** double-click the .jar file to start it, but double-click the supplied shell script **start.sh**. This will make 1.5Gb memory available to the application.

JAVA

To use RockerBox, you will need Java 1.6 or newer. The Windows executable will point you to the download if it's not installed, otherwise get it here:

<http://www.java.com>

PERCOLATOR

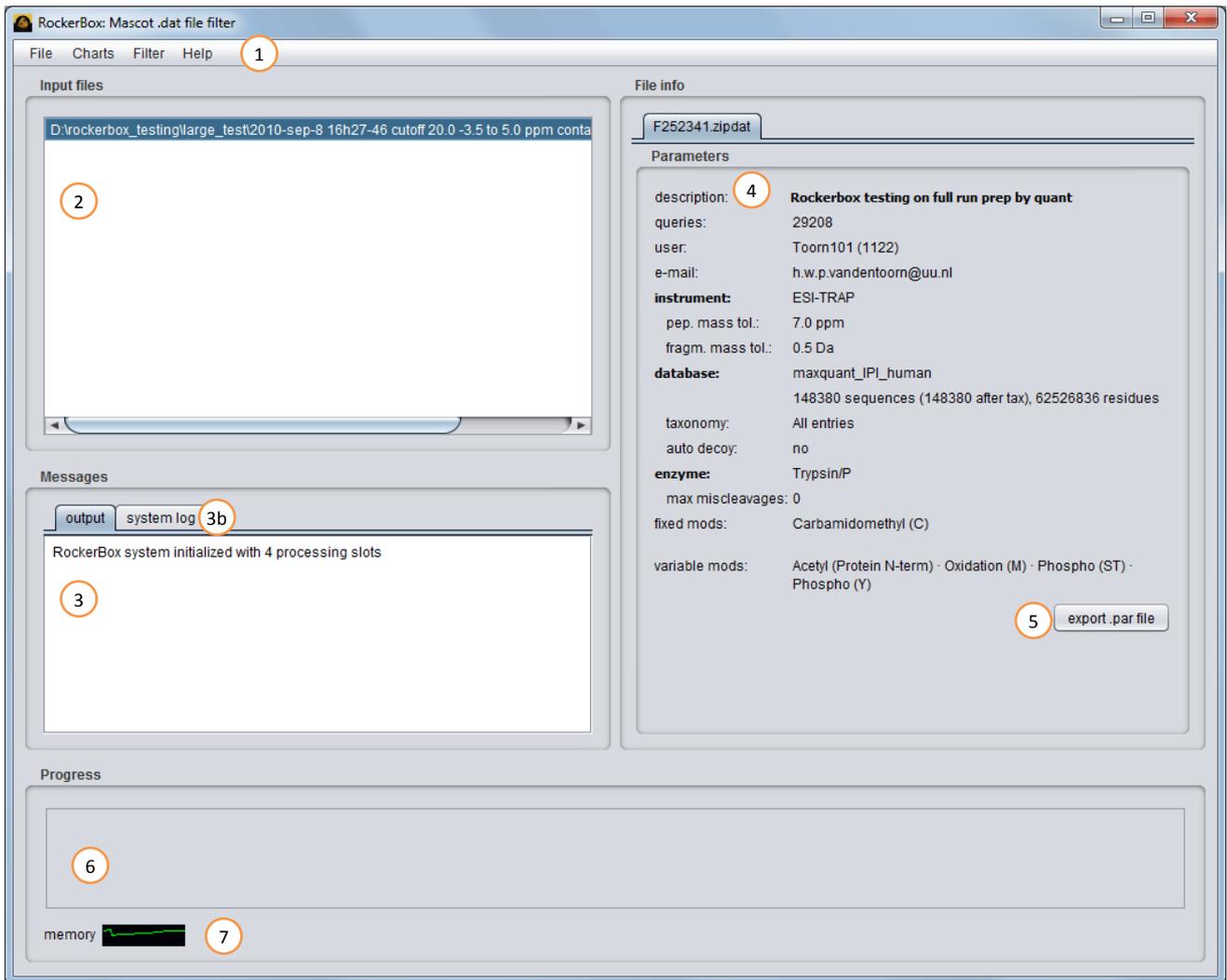
If you're interested in using the Percolator algorithm for filtering, get the executable here:

<http://github.com/percolator/percolator/downloads>

Product names, logos, brands, and other trademarks featured or referred to within these documents are the property of their respective trademark holders.

These trademark holders are not affiliated with Utrecht University, the Biomolecular Mass Spectrometry and Proteomics Group, the Netherlands Proteomics Centre nor the Netherlands Bioinformatics Centre. No endorsement of aforementioned groups or our products is expressed or implied by the mention of such trademarks.

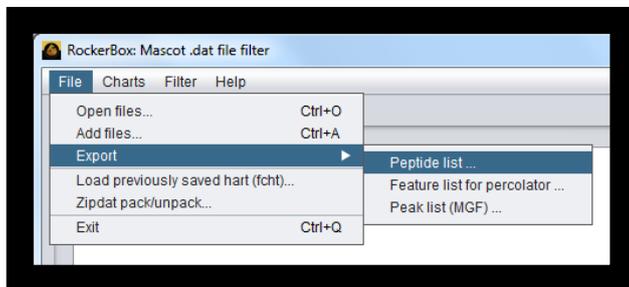
ROCKERBOX MAIN WINDOW OVERVIEW



- 1 The menu area which gives access to all functionality. The menus are explained later.
- 2 The file list, showing the files that are loaded to be processed. Multiple files can be listed here, enabling batch processing of files.
- 3 Message area, showing progress on the current operation. 3b System log tab, showing any errors if they occur. Here, you see that the application has started successfully and the number of processing slots that have been annotated. The number of slots depends on the number of processors your system reports and indicate the number of simultaneous operations RockerBox can perform.
- 4 Information area, showing search parameters for the currently selected file.
- 5 Button to export a parameter (.par) file for Mascot Daemon
- 6 Area for progress bar(s). If multiple processes run simultaneously, more than one progress bar may be visible.
- 7 Memory meter, indicating memory usage of the application within its set limits. Clicking on it will force the release of unused memory.

EXPORTING DATA

If you need to get data from your .dat file there are several options in RockerBox, found under the 'Export' menu in the File main menu.



EXPORT PEPTIDE LIST

Exporting a Peptide list will give you a tab delimited file with information about all *first ranking* PSMs, including:

Row name	
1	mascot query number
2	scan number(s)*
3	retention time*
4	raw file*
5	sequence
6	Score
7	peptide mr
8	mass delta (abs)
9	mass delta (ppm)
10	modifications
11	all protein matches

* Only shown when this information can be extracted from the .dat file.

EXPORT FEATURE LIST FOR PERCOLATOR

The **Percolator based filter** generates a text file with features for every highest-ranking PSM in the .dat file. Because this information might be useful outside of Percolator, or if you wish to run the Percolator algorithm separately from RockerBox, this export method is available. The features include:

Feature	Description
1	id Identifier*
2	label -1 if decoy, 1 if target PSM
3	charge Precursor charge
4	score Mascot score
5	deltaScore Difference between current rank score and 'next' rank score
6	mr Measured precursor mass
7	deltaM Delta mass between precursor mass and matched peptide mass
8	deltaMPpm deltaM relative to matched peptide mass
9	absDeltaM Absolute value of deltaM
10	absDeltaMPpm Absolute value of deltaMPpm
11	isoDeltaM Delta mass allowing for 1, 2, 3 or 4 Dalton difference
12	isoDeltaMPpm isoDeltaM relative to matched peptide mass
13	missedCleavages Number of missed cleavages
14	fragMassError* RMS error of the MS2 spectrum to the theoretical spectrum
15	totalIntensity* Total intensity of the MS2 spectrum
16	intMatchedTot* Total intensity of matched MS2 peaks
17	relIntMatchedTot* intMatchedTot divided by totalIntensity
18	fracIonsMatched* Fraction of all MS2 peaks matched
19	peptide Peptide sequence
20	proteins The list of proteins from the search database that contain the peptide sequence

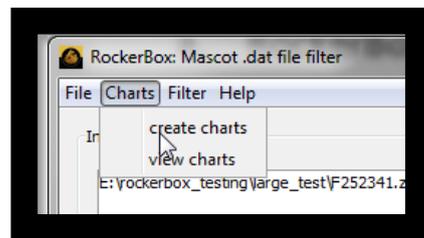
*RockerBox uses the form **db*_querynumber_rank*, in which **db** may be 'target' for Mascot automatic decoy real database, 'decoy' for Mascot automatic decoy scrambled database or 'combined' for a concatenated decoy strategy

EXPORT PEAK LIST (MGF)

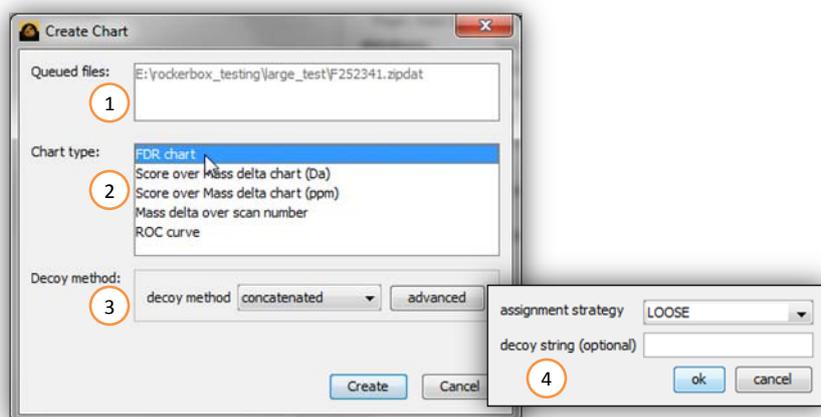
It is possible to extract the peak lists from the Mascot output file. Only spectra that Mascot has matched to a peptide can be exported, since for the remainder no precursor mass is present in the .dat file.

CREATING CHARTS

To get an overview of the data in your file, quickly create charts with this tool.



DIALOG



- 1 A list of files is shown in the 'queued files' list for which charts will be created.
- 2 In the chart type list, a chart type is selected.
- 3 All charts can be made using data from either 'concatenated' database searches or from a 'automatic decoy' search as performed by Mascot.
- 4 For the 'concatenated' option, the 'advanced' button becomes active, which gives the option to enter a string for recognition of decoy proteins (e.g. "REV_") if this string is not present in the current implementation of RockerBox.

Decoy assignment may be ambiguous when the same sequence matches both (a number of) 'target' and 'decoy' database entries. In strict mode, a PTM is assigned to be a decoy more easily than in loose mode, slightly changing FDR estimations. Options:

STRICT: label a peptide as 'decoy' if *any* of the matching proteins come from the decoy database

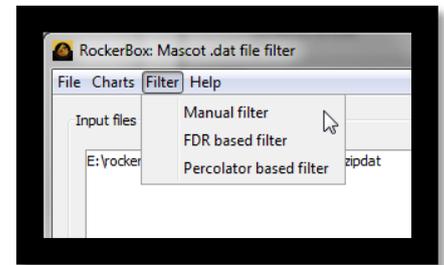
LOOSE: label the peptide as 'decoy' if *all* of the matching proteins come from the decoy database

CHART OUTPUT

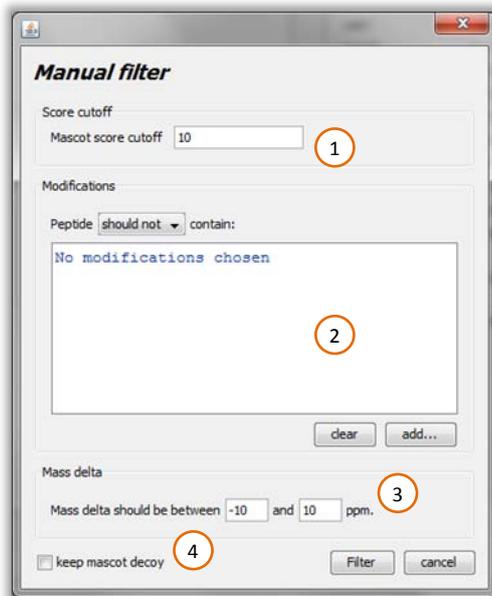
After a chart is finished, a thumbnail image will appear in a separate dialog. Double-clicking on a thumbnail will show an interactive window. Dragging the mouse diagonally from left-top to right-down inside this window allows you to zoom into the chart. A pop-up menu is upon right-click, containing the export functionality (save as...) and fine control over axes and titles from the properties location.

FILTERING

There are currently three ways to filter a .dat file, to create a new and smaller .dat file: a manual filter, an FDR based filter and a percolator based filter.



MANUAL FILTER

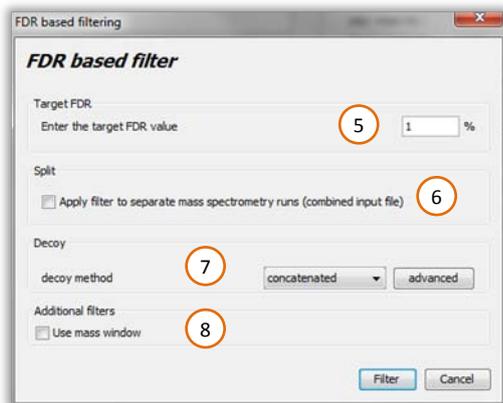


There are four criteria to filter PSMs from a .dat file:

- 1 The mascot score cutoff is the minimum mascot ions score that a PSM should have to pass filtering.
- 2 The modifications chooser allow for the choice of one or more modifications and whether to include or exclude the chosen modifications.
- 3 Precursor masses are filtered according the mass window entered here.
- 4 When filtering based a separate decoy database strategy, you might want to keep the the Mascot automatic decoy search results for later analysis. Both database searches will be filtered according to the criteria set in this filter dialog.

As a side effect, decoy PSMs from the 'concatenated' search method are removed. This filter considers PSMs from all ranks, re-ranking the PSMs to keep the rank order.

FDR BASED FILTER



- 5 A target FDR value is entered here (in percentage).
- 6 Using the Split option, mass spectrometry runs inside the .dat file can be filtered separately, or all together. The result of the filtering is always a single .dat file.
- 7 The algorithm needs input about the type of decoy used for the experiment. The function of the 'advanced' button is described in the **Creating Charts** section.
- 8 An additional filter can be added to the FDR based filter to restrict FDR calculations and filtering to a mass window. Clicking this option box will reveal a mass delta input field similar to the one in the manual filter dialog 3.

As a side effect, decoy PSMs from the 'concatenated' search method are removed. This filter considers PSMs from all ranks, re-ranking the PSMs to keep the rank order.

PERCOLATOR BASED FILTER



9, 10 and 11 correspond to 5, 6 and 7 of the FDR based filter dialog.

12 In the score output pull-down menu, the output score can be chosen. Choosing Mascot Score retains the Mascot ions score found from the input file. Choosing $-10 \log(q\text{-value})$ or $-10 \log(\text{PEP})$ replaces the scores in the output file with output values from the Percolator algorithm.

13 The location of the percolator executable, should be set here before this filter will work.

TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIBUTION

Apache License
Version 2.0, January 2004
<http://www.apache.org/licenses/>

1. DEFINITIONS.

"License" shall mean the terms and conditions for use, reproduction, and distribution as defined by Sections 1 through 9 of this document.

"Licensor" shall mean the copyright owner or entity authorized by the copyright owner that is granting the License.

"Legal Entity" shall mean the union of the acting entity and all other entities that control, are controlled by, or are under common control with that entity. For the purposes of this definition, "control" means (i) the power, direct or indirect, to cause the direction or management of such entity, whether by contract or otherwise, or (ii) ownership of fifty percent (50%) or more of the outstanding shares, or (iii) beneficial ownership of such entity.

"You" (or "Your") shall mean an individual or Legal Entity exercising permissions granted by this License.

"Source" form shall mean the preferred form for making modifications, including but not limited to software source code, documentation source, and configuration files.

"Object" form shall mean any form resulting from mechanical transformation or translation of a Source form, including but not limited to compiled object code, generated documentation, and conversions to other media types.

"Work" shall mean the work of authorship, whether in Source or Object form, made available under the License, as indicated by a copyright notice that is included in or attached to the work (an example is provided in the Appendix below).

"Derivative Works" shall mean any work, whether in Source or Object form, that is based on (or derived from) the Work and for which the editorial revisions, annotations, elaborations, or other modifications represent, as a whole, an original work of authorship. For the purposes of this License, Derivative Works shall not include works that remain separable from, or merely link (or bind by name) to the interfaces of, the Work and Derivative Works thereof.

"Contribution" shall mean any work of authorship, including the original version of the Work and any modifications or additions to that Work or Derivative Works thereof, that is intentionally submitted to Licensor for inclusion in the Work by the copyright owner or by an individual or Legal Entity authorized to submit on behalf of the copyright owner. For the purposes of this definition, "submitted" means any form of electronic, verbal, or written communication sent to the Licensor or its representatives, including but not limited to communication on electronic mailing lists, source code control systems, and issue tracking systems that are managed by, or on behalf of, the Licensor for the purpose of discussing and improving the Work, but excluding communication that is conspicuously marked or otherwise designated in writing by the copyright owner as "Not a Contribution."

"Contributor" shall mean Licensor and any individual or Legal Entity on behalf of whom a Contribution has been received by Licensor and subsequently incorporated within the Work.

2. GRANT OF COPYRIGHT LICENSE.

Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable copyright license to reproduce, prepare Derivative Works of, publicly display, publicly perform, sublicense, and distribute the Work and such Derivative Works in Source or Object form.

3. GRANT OF PATENT LICENSE.

Subject to the terms and conditions of this License, each Contributor hereby grants to You a perpetual, worldwide, non-exclusive, no-charge, royalty-free, irrevocable (except as stated in this section) patent license to make, have made, use, offer to sell, sell, import, and otherwise transfer the Work, where such license applies only to those patent claims licensable by such Contributor that are necessarily infringed by their Contribution(s) alone or by combination of their Contribution(s) with the Work to which such Contribution(s) was submitted. If You institute patent litigation against any entity (including a cross-claim or counterclaim in a lawsuit) alleging that the Work or a Contribution incorporated within the Work constitutes direct or contributory patent infringement, then any patent licenses granted to You under this License for that Work shall terminate as of the date such litigation is filed.

4. REDISTRIBUTION.

You may reproduce and distribute copies of the Work or Derivative Works thereof in any medium, with or without modifications, and in Source or Object form, provided that You meet the following conditions:

- a. You must give any other recipients of the Work or Derivative Works a copy of this License; and
- b. You must cause any modified files to carry prominent notices stating that You changed the files; and
- c. You must retain, in the Source form of any Derivative Works that You distribute, all copyright, patent, trademark, and attribution notices from the Source form of the Work, excluding those notices that do not pertain to any part of the Derivative Works; and
- d. If the Work includes a "NOTICE" text file as part of its distribution, then any Derivative Works that You distribute must include a readable copy of the attribution notices contained within such NOTICE file, excluding those notices that do not pertain to any part of the Derivative Works, in at least one of the following places: within a NOTICE text file distributed as part of the Derivative Works; within the Source form or documentation, if provided along with the Derivative Works; or, within a display generated by the Derivative Works, if and wherever such third-party notices normally appear. The contents of the NOTICE file are for informational purposes only and do not modify the License. You may add Your own attribution notices within Derivative Works that You distribute, alongside or as an addendum to the NOTICE text from the Work, provided that such additional attribution notices cannot be construed as modifying the License.

You may add Your own copyright statement to Your modifications and may provide additional or different license terms and conditions for use, reproduction, or distribution of Your modifications, or for any such Derivative Works as a whole, provided Your use, reproduction, and distribution of the Work otherwise complies with the conditions stated in this License.

5. SUBMISSION OF CONTRIBUTIONS.

Unless You explicitly state otherwise, any Contribution intentionally submitted for inclusion in the Work by You to the Licensor shall be under the terms and conditions of this License, without any additional terms or conditions. Notwithstanding the above, nothing herein shall supersede or modify the terms of any separate license agreement you may have executed with Licensor regarding such Contributions.

6. TRADEMARKS.

This License does not grant permission to use the trade names, trademarks, service marks, or product names of the Licensor, except as required for reasonable and customary use in describing the origin of the Work and reproducing the content of the NOTICE file.

7. DISCLAIMER OF WARRANTY.

Unless required by applicable law or agreed to in writing, Licensor provides the Work (and each Contributor provides its Contributions) on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied, including, without limitation, any warranties or conditions of TITLE, NON-INFRINGEMENT, MERCHANTABILITY, or FITNESS FOR A PARTICULAR PURPOSE. You are solely responsible for determining the appropriateness of using or redistributing the Work and assume any risks associated with Your exercise of permissions under this License.

8. LIMITATION OF LIABILITY.

In no event and under no legal theory, whether in tort (including negligence), contract, or otherwise, unless required by applicable law (such as deliberate and grossly negligent acts) or agreed to in writing, shall any Contributor be liable to You for damages, including any direct, indirect, special, incidental, or consequential damages of any character arising as a result of this License or out of the use or inability to use the Work (including but not limited to damages for loss of goodwill, work stoppage, computer failure or malfunction, or any and all other commercial damages or losses), even if such Contributor has been advised of the possibility of such damages.

9. ACCEPTING WARRANTY OR ADDITIONAL LIABILITY.

While redistributing the Work or Derivative Works thereof, You may choose to offer, and charge a fee for, acceptance of support, warranty, indemnity, or other liability obligations and/or rights consistent with this License. However, in accepting such obligations, You may act only on Your own behalf and on Your sole responsibility, not on behalf of any other Contributor, and only if You agree to indemnify, defend, and hold each Contributor harmless for any liability incurred by, or claims asserted against, such Contributor by reason of your accepting any such warranty or additional liability.

END OF TERMS AND CONDITIONS