

By selecting the "accept licence agreement" (or any equivalent) button and/or by using, copying or distributing this Software or any portion thereof, YOU (the "User") ACCEPT ALL TERMS AND CONDITIONS OF THIS LICENCE, including in particular the limitations on use, transferability, warranty and liability. The following terms and conditions are enforceable against you and any legal entity that obtained the Software and on whose behalf it is used. If you are agreeing to these terms on behalf of a company or other legal entity, you represent that you have the legal authority to bind that company or legal entity to these terms. IF YOU DO NOT HAVE SUCH AUTHORITY OR IF YOU DO NOT WISH TO BE BOUND TO THESE TERMS DO NOT USE THIS SOFTWARE.

*The European Union (hereinafter "the Licensor") is the owner of the copyright and other intellectual and industrial property rights, trade secrets, and know-how related to the software JRC-Eurovoc-Indexer (JEX), over which it has the power of disposal regardless geographical or other limitations.*

*JEX is a software product of the Joint Research Centre of the European Commission which automatically assigns a ranked lists of descriptors (classes) from the Eurovoc thesaurus (<http://eurovoc.europa.eu/>) to a written plain-text input file. The tool can train itself on document collections to which descriptors have previously been manually assigned. The JRC has trained the tool on the 23 official EU languages, but it can in principle be trained for many more languages if manually indexed document collections are available, and it can be trained for descriptors of thesauri other than Eurovoc. The main parts of the tool are the indexing component and a self-learning component, but it also includes a pre-processing and a post-processing component, as well as an interactive user interface that allows users to manually view and correct the automatic assignment.*

*The main purpose of the software is to complement or replace the human intellectual indexing process of classifying documents according to the Eurovoc thesaurus. However, due to the nature of the multilingual Eurovoc thesaurus (each subject domain class is represented by a language-independent numerical identifier), the software can also serve to represent texts in a language-independent way in order to allow the comparison of documents across languages.*

## 1. DEFINITIONS

.1. The under mentioned terms printed with an initial capital letter shall have herein the following meanings unless the context otherwise requires:

"Computer" shall mean an electronic device that accepts information in digital or similar form and manipulates it for specific results based on a sequence of instructions.

"Effective Date" shall mean the date of first Use by the User.

"Parties" shall jointly refer to both the Licensor and the User in this licence while "Party" shall refer to either of the Parties as the context provides;

"Software" shall mean any series of instructions constituting a computer-executable program or programs as well as any modification or updated versions of the computer software. The Software shall also include the whole content of the files (provided either by electronic download, on physical media or any other method of distribution), disk(s), CD-ROM(s) or other media being object of this licence and any database and support materials licensed by the Licensor including - but not limited to - manuals, flow charts and specifications relating to the Software as well, as described in Annex A.

"Use" or "Used" in connection with the Software, as defined below, shall mean storing any portion of the Software in a machine, and/or transmitting any portion of the Software to a machine for processing and/or compiling, executing or interpreting any machine instructions contained in the Software, and/or displaying any portion of the Software in connection with the processing of such machine instructions.

## 2. GRANT OF RIGHTS AND LIMITATIONS

.1. Subject to compliance with the terms of this licence, the Licensor grants the User a personal, non-exclusive and royalty-free licence to Use the Software.

.2. The User may install, access and Use the Software within the computer environment of the User, including internal network deployments and/or several separate standalone deployments. The User may make copies of the Software, solely in machine readable form for back-up and archival purpose, provided that User reproduces on this copy all the proprietary notices which appear on, with or in the Software.

.3. The User may not rent, lease, sublicense, assign, transfer or grant any kind of rights regarding the Software or any portions thereof in any form to any third party except as may be expressly permitted herein. You may, however transfer all your rights to Use the Software to a third party under the strict condition that you also transfer this licence and the Software.

.4. The User may not modify, reverse engineer, decompile, disassemble, create derivative works based on, or copy the Software or any part of the Software, except otherwise allowed by law or by the present licence.

.5. The User may not remove or alter any Software identification, proprietary notices, labels or trademarks which appear on or in the Software.

.6. The Licensor shall have no obligation for the installation, technical support and maintenance of the Software.

### 3. OWNERSHIP AND LICENCE

.1. The Software is not sold to User who shall not acquire any right, title or interest (including without limitation copyright or other right in the nature of copyright or any other intellectual property right whatsoever) in the Software, which shall remain the sole property of the Licensor.

.2. Any copy or partial copy of the Software shall be owned by the Licensor subject to the license rights granted to the User and shall be considered as being regulated by this licence.

.3. The User's rights to Use the Software are specified in this licence, and the Licensor retains all rights not expressly granted to User in this licence. No right or licence under any patent application, issued patent, know-how or other proprietary information is granted or shall be granted by implication unless provided herein.

### 4. DURATION AND TERMINATION

.1. This licence is granted on a perpetual basis and shall be effective as of the Effective Date. However, Licensor and User may agree to terminate this licence at anytime without cause.

.2. This licence shall automatically terminate with an immediate effect if the User is in breach of this licence. This termination shall not prevent the Licensor from claiming any further damages. Upon termination for any reason, the User shall destroy or return the Software to the Licensor and any copy made - partial or whole - in its possession. Upon termination for any reason and subject to a written request from the Licensor, the User shall certify by means of a written document duly signed by a legal representative that the provisions of the present Article have been respected, and this within a time limit of fifteen (15) calendar days as from the date of the request.

.3. The termination shall not relieve the User from its liability to respect all the obligations claimable before the termination date. In particular, the provisions of the obligations relating to the performance, the disclaimer of guarantees and warranties and the indemnification and limitations of liabilities shall survive the termination of this licence, howsoever caused, but this shall not imply or create any continued right to Use the Software after termination of this licence.

### 5. INTELLECTUAL PROPERTY RIGHTS

.1. The name, the copyright and the intellectual and industrial property rights related to the Software, including database rights where enforceable, are the exclusive property of the Licensor.

.2. "JRC" is a Licensor's registered trademark. The User undertakes not to file trademark applications containing the denomination of the Software or parts of such denomination in their own name.

.3. Third-party proprietary software necessary for the functioning of the Software may be distributed in bundle with the Software. Such distribution and use are strictly subject to the terms and conditions set forth in the respective proprietary licences, as provided in Annex A, to which the User accepts to be bound.

### 6. GUARANTEES, WARRANTY AND LIMITATIONS OF LIABILITIES

.1. The Software is provided "as is" without warranty of any kind, either express or implied, including, but not limited to, any implied warranty against infringement of third parties' property rights, of merchantability, integration, satisfactory quality and fitness for a particular purpose. In particular, the software produces results automatically and its output may therefore contain errors. The Licensor is not responsible for any wrongly produced output.

.2. The entire risk as to the Use, quality, and performance of the Software is with the User. The Licensor will not be liable for any incidental, consequential, direct or indirect damages including but not limited to the loss of data, lost profits, or any other financial loss arising from the Use of, or inability to Use, even if the Licensor has been notified of the possibility of such loss, damages, claims or costs or for any claim by any third party.

7. SEVERABILITY

.1. If any term, provision, covenant or condition of this licence is held or becomes invalid or unenforceable for any reason, the Parties agree to give the remaining provisions the maximum practicable enforceability that the applicable law allows, unless such remaining provisions would frustrate the intent of the Parties, in which case the licence will immediately terminate.

8. APPLICABLE LAW AND LEGAL VENUE

.1. This licence shall be governed by the law of the European Union and, where necessary, by the substantive law of Belgium.

.2. In case of any dispute or difference between the Parties arising out of or in connection with this licence, the Parties shall settle it by mutual agreement. Such effort shall be deemed to have failed when one of the Parties so notifies the other in writing. In that case, each Party may initiate proceedings before the General Court of the European Union in Luxembourg.

ANNEX A – TECHNICAL SPECIFICATIONS OF THE SOFTWARE

Version: 1.0

**Language:** The resource and software are not limited to any specific natural language. It is inherently multilingual. The software has been pre-trained for the 23 official languages of the European Union. The trained software is distributed separately for each of the languages so that users can use the software for one or more languages of their choice. The documentation is written in English.

**Components:** The software consists of six components: an interactive graphical user interface (GUI); a pre-processing tool that can help to convert certain input text files to the format needed by the main indexing component; the main indexing component, which assigns a ranked list of Eurovoc descriptors; a post-processing component, which may enrich the output of the main component with additional information from the Eurovoc thesaurus, as well as with meta-information from the indexing process; a training component, which can be used to train the classification software with user-provided manually indexed documents; and a component to test the result of the automatic training. The software is accompanied by numerous configuration files, by a copy of the Eurovoc thesaurus, and by collections of manually Eurovoc-indexed training documents in the software-specific ‘compact format’.

**Functionalities:** The indexing software can take plain-text files as input and assign ranked lists of Eurovoc descriptors to each of these texts. The graphical user interface allows to view the assignment results of the indexing step, including high-lighting the words that caused the software to assign individual descriptors. The GUI furthermore allows to manually moderate the results, by deleting automatically assigned descriptors, by searching the Eurovoc thesaurus for other descriptors and by adding them to the list of assigned descriptors. Finally, it allows to save the results in XML format. The training software allows to learn new classifiers by analysing user-provided, manually indexed document collections and thus to improve the indexing result by providing more and better training collections. Also, it allows to re-train the software with different - and possibly better - parameter settings. The pre-processing component can optionally be used to convert the input files into the format needed by the main indexing component. The post-processing component can optionally be used to enrich the output of the main indexing component with additional information about the Eurovoc thesaurus and about the indexing process. The training component can re-train the software with user-provided thesaurus-indexed documents.

**Documentation:** The software is accompanied by documentation in HTML and in PDF format. Additional scientific-technical details regarding the indexing algorithm and the evaluation of the software are available in the publication shown below:

Pouliquen Bruno, Steinberger Ralf, Camelia Ignat (2003). Automatic Annotation of Multilingual Text Collections with a Conceptual Thesaurus. In: Proceedings of the Workshop *Ontologies and Information Extraction* at the Summer School *The Semantic Web and Language Technology - Its Potential and Practicalities* (EUROLAN'2003). Bucharest, Romania, 28 July - 8 August 2003.

Further and newer publications on amendments of the algorithm will become available on: [http://langtech.jrc.ec.europa.eu/JRC\\_Publications.html](http://langtech.jrc.ec.europa.eu/JRC_Publications.html). Programming language: The software is programmed in Java.