



ACCURAT

Analysis and Evaluation of Comparable Corpora
for Under Resourced Areas of Machine Translation

www accurat-project.eu

Project no. 248347



Deliverable D6.7
Report on IPR of the project results

Version No. 1.0
31/12/2011

Document Information

Deliverable number:	D6.7
Deliverable title:	Report on IPR of the project results
Due date of deliverable:	31/12/2011
Actual submission date of deliverable:	31/12/2010
Main Author(s):	Marko Tadić
Participants:	All Partners
Internal reviewer:	Tilde
Workpackage:	WP6
Workpackage title:	Dissemination
Workpackage leader:	FFZG
Dissemination Level:	1.0 : PU
Version:	V0.5
Keywords:	IPR issues, software, resources

History of Versions

Version	Date	Status	Name of the Author (Partner)	Contributions	Description/ Approval Level
V0.5	28/12/2010	Completed	FFZG	Draft	Submitted to Tilde
V1.0	30/12/2011	Final	Tilde	Final modifications	Submitted to PO

EXECUTIVE SUMMARY

This report on IPR issues gives recommendations how to treat IPR of ACCURAT results. It examines and describes existing licensing schemes that could be applicable for ACCURAT results, particularly ones proposed by META-NET, i.e., META-SHARE Commons licence, and analyses the licensing mechanisms of third party tools used in development of ACCURAT tools. At the end this deliverable give several recommendations how to formulate licences of ACCURAT results by the end of the project.

Table of Contents

Abbreviations	4
1 Introduction	5
2 Consulted licensing schemes applicable to ACCURAT results	6
2.1 GNU General Public License	6
2.2 Creative Commons (CC)	6
2.3 Apache 2.0	7
2.4 META-SHARE	7
3 Commitments relevant for IPR in ACCURAT DoW	9
4 META-NET IPR recommendations	10
5 Third party licensing information	16
6 Recommended licensing schemes	22
6.1 META-SHARE users	22
6.2 Non-META-SHARE users	22
7 Conclusion	24

Abbreviations

Table 1 Abbreviations

Abbreviation	Term/definition
ACCURAT	Analysis and Evaluation of Comparable Corpora for Under Resourced Areas of Machine Translation
ASF	The Apache Software Foundation
BY	Done by
CC	Creative Commons
DoW	Description of Work
ELRA	European Language Resources Association
GNU	“GNU” is a recursive acronym for “GNU's Not Unix!”
GPL	General Public License
HLT	Human Language Technologies
ICT	Information and communication technologies
ICT-PSP	ICT Policy Support Programme
IPR	Intellectual property rights
KEA	Kayphrase extraction algorithm
LGPL	Lesser General Public License
LR	Language resources
LR&T	Language resources and tools
LT	Language technologies
META-NET	Multilingual Europe Technology Alliance Network of Excellence
MIT	Massachusetts Institute of Technology
MSC	META-SHARE Commons licence
Ms-RSL	The Microsoft Reference Source License
NC	Non-commercial
ND	No derivatives
SA	Share alike

1 Introduction

It is generally accepted that the IPR treatment of a single software tool could be problematic by itself. In the case of ACCURAT project results, we have the whole toolkit, i.e., set of complex and mutually intertwined pieces of software that rely not just on each other, but also on third party tools and libraries. These third party tools and libraries use different licensing mechanisms thus restricting to a certain extent the general idea ACCURAT partners had from the beginning – to produce a publicly available, open source set of tools for processing comparable corpora.

In order to solve this complicated network of IPR relations, we we had to consult many existing licensing schemas that would be applicable to ACCURAT results.

2 Consulted licensing schemes applicable to ACCURAT results

There are several possible decisions IPR owners could choose between when it comes to the legal treatment of their IPR while LR&T are being used:

- a) waive all rights (copyleft, no right reserved, public domain)
 - this is not allowed with all jurisdictions (e.g. Moral rights under French law)
- b) apply the copyright law
 - different regulations in different jurisdictions
 - each new user should ask for prior agreement from the consortium (letter of intent, bi-/multi-lateral contract)
 - such agreement can be customized for every user
- c) license the tools and data and make such license publicly available for all potential users
 - permissive and implicit license such as CC, GPL or similar
 - license with some restrictions regarding the redistribution (e.g. ELRA or LDC)
 - in-house license, regulates granted rights on case by case

We have consulted a number of possible licensing mechanisms, that are used for distribution of digital data and programs, and collected their most important features relevant for distribution of ACCURAT results.

2.1 GNU General Public License

GNU GPL v3.0 (<http://www.gnu.org/licenses/gpl-3.0.txt>) is the most widely used free software licence. Its main limitation is the obligation that if any software under this license is being included into a more complex software package, the whole package has to be issued under GNU GPL also. This is in accordance with the copyleft principle that GPL licensing mechanism uses.

The derivative of GNU GPL is GNU LGPL that is mainly used for licensing programming libraries and runtime environments. It has more relaxed copyleft restrictions than GNU GPL license.

We have considered these licenses because most of academic partners in ACCURAT project expressed their readiness to publish the tools under that or similar open source license.

However, GNU GPL is primarily oriented towards programs and is not very useful for distribution of language resources, i.e., corpora, that are collected within ACCURAT project.

It should also be noted that the copyleft applies only to the software and not to its output. For example, a tagger running as a modified derivative of a GPL'ed tagging system is not required to distribute its changes to the underlying software because its output is not a derivative.

2.2 Creative Commons (CC)

There are several possible CC v3.0 (<http://creativecommons.org>) licensing conditions that can be combined into a licence. CC licences can be non-restrictive or restrictive:

- non-restrictive
 - CCo (CC Zero)
 - no limitations at all

- restrictive
 - CC BY
 - attribution needed, i.e. “done by” (BY)
 - CC BY-ND
 - attribution needed
 - no derivatives (ND)
 - CC BY-SA
 - attribution needed
 - share alike, i.e. all derivatives have the same license (SA)
 - CC BY-NC
 - attribution needed
 - non-commercial (NC)
 - CC BY-NC-SA
 - attribution needed
 - non-commercial (NC)
 - share alike (SA)
 - CC BY-NC-ND
 - attribution needed
 - non-commercial (NC)
 - no derivatives (ND)

CC has quite elaborated system for building licence schemas with different conditions as building blocks where each condition imposes specific restriction of usage. CC is very well known internationally recognised and impartial licensing system that is oriented primarily towards the digital content accessible over Internet. It gives the right balance of freedom and desired level of restrictiveness and the combinations is open for choice by owner. For this reasons it remains a good candidate for ACCURAT results licensing mechanism. Also, it is much better suited for distribution of language resources since it is oriented towards any kind of digital content.

2.3 Apache 2.0

Apache 2.0 (<http://www.apache.org/licenses/LICENSE-2.0.html>) is similar to GNU GPL, but without limitations imposed by copyleft conditions. It is

- freely changeable and distributable;
- requires only inclusion of attribution (like CC BY)
- it allows commercial and non-commercial use

In this respect Apache 2.0 license could be considered as a good candidate for distribution of ACCURAT tools. However, for distribution of ACCURAT language resources Apache 2.0 licence is not well suited.

2.4 META-SHARE

The META-NET deliverable D6.1.1 *META-SHARE: Licenses, Legal, IPR and Licensing issues* (version from March 2011) defines the so called META-SHARE Commons licence (MSC):

- used for licensing within the META-SHARE framework

- allows sharing, re-using and remixing LR&T inside the framework
- it allows commercial and non-commercial use
- similar to CC but limitation that LR&T will be available to META-SHARE members

The META-SHARE Commons licence is highly relevant for ACCURAT project results distribution because of a series of commitments ACCURAT partners made already in the time of project proposal.

3 Commitments relevant for IPR in ACCURAT DoW

In ACCURAT DoW we have used the acronym T4ME since it was an official one at the time of writing ACCURAT project proposal. In the meantime with the inclusion of three additional ICT-PSP projects this effort turned into the META-NET alliance. In this report and as particularly in quotes from ACCURAT DoW, the T4ME and META-NET should be considered synonymous.

In ACCURAT DoW we made a clear commitment of collaboration with META-NET (T4ME) on page 115:

Of particular interest in [corr. is] the collaboration between ACCURAT and T4ME in the framework of Pillar 2 “Building an Open Resource Infrastructure” which sets out to create an open, integrated, secured, and interoperable language resources (LR) and language technologies (LT) infrastructure for the HLT (Human Language Technologies) domain.

There (page 115) we stated also that:

T4ME will offer a wide range of services including: registration, authorization/authentication, (...) and access/distribution services including IPR and legal clearance issues.

and:

Given the objectives and expected outputs of ACCURAT and the mandate of T4ME, the following collaboration is planned:

- *Publicizing, documenting and sharing of language resources collected by ACCURAT.*
- *Publicizing, documenting and sharing of ACCURAT open-source tools. In the case of sharing, such tools can be offered in a downloadable mode or as web services.*

while on page 116 we stated also:

T4ME language resource infrastructure will be used as a primary channel for distribution of tools and resources developed by ACCURAT.

These statements represent a firm decision that ACCURAT project will distribute its results through META-NET’s distributional platform that emerged in the mean time and is named META-SHARE. This also implies that ACCURAT project results could use IPR and clearance services that META-SHARE would provide to its members. This also implies that the IPR mechanisms for ACCURAT results are expected to be compliant to META-SHARE licensing mechanisms, if not completely, then to a large extent, to avoid incompatibilities. In this respect we had to consult META-SHARE licensing mechanisms and try to apply them to ACCURAT project results although these mechanisms should not be considered as a job completely finished since the META-SHARE licensing schemes are still not firmly defined. Also, we have to respect the licensing limitations of third-parties whose tools, libraries and resources were used to produce ACCURAT resources and tools.

4 META-NET IPR recommendations

The META-NET deliverable D6.1.1 *META-SHARE: Licenses, Legal, IPR and Licensing issues* (version from March 2011) gives an exhaustive insight into the field of LRs and LTs, describes their production process and its legal implications, and their usage and its legal implications. It also presents the current European legal frameworks with respect to LRs. Within this framework different licensing legal frameworks are being discussed which are of particular importance for ACCURAT.

After considerable reflections on different licensing mechanisms, in D6.1.1 META-NET has suggested its own **META-SHARE Commons licence** that has been inspired by Creative Commons licensing mechanism. The text of this licence¹ follows.

The META-SHARE Commons licence

META-SHARE IS NOT A LAW FIRM AND DOES NOT PROVIDE LEGAL SERVICES. DISTRIBUTION OF THIS LICENCE DOES NOT CREATE AN AGENT-CLIENT RELATIONSHIP. META-SHARE PROVIDES THIS INFORMATION ON AN "AS-IS" BASIS. METASHARE MAKES NO WARRANTIES REGARDING THE INFORMATION PROVIDED, AND DISCLAIMS LIABILITY FOR DAMAGES RESULTING FROM ITS USE.

Licence

This META-SHARE public licence enables You to Use the Resource worldwide for noncommercial purposes provided that You keep to the terms of this Licence.

1. Definitions of Capitalised Words

"Collective Work" means a work made up of the Work in its entirety in unmodified form, along with other work, assembled into a collective whole.

"Derivative" means any material that Uses the Resource (or any substantial part of it) in any material form whatsoever (such as a translation, dramatisation or abridgment), other than (i) as a whole and in unmodified form or (ii) by modifying it as may be technically necessary to Use it in a different mode or format; but a Collective Work is not a Derivative for the purpose of this Licence.

"Derogatory Treatment" means a treatment which distorts or mutilates the Resource or is otherwise prejudicial to the honour or reputation of the Original Author or the Licensor.

"Licence Elements" means the following licence attributes indicated in the title of this Licence: Attribution, Non-commercial, Share-Alike.

"Licensor" means the Person offering the Resource under the terms and conditions of this Licence.

¹ Although everywhere in the text of the META-SHARE Commons licence the term "resource" is used it clearly implies "language resource" and "language tool", thus allowing that this licence is applicable to both, resources and tools. This is important for ACCURAT project since tools represents a considerable part of project results.

"Original Author" means the Person who obtained any copyright in the Resource or any parts of it and is named as such in the Attribution Data.

"Person" means a natural person or a body of persons corporate or incorporate.

"Use", as a verb, means doing any act which is restricted by copyright or neighboring rights (including database rights), whether in the original medium or any other; and includes modifying the Resource as may be technically necessary to Use it in a different mode or format.

"Work" means any work protected by copyright (or by database rights if applicable) which is offered under the terms of this Licence, and includes works forming only a part of the Resource as well works as incorporated in any Collective Work.

"You" means the Person acquiring rights under this Licence.

"Attribution Data" means a field of metadata accompanying every METASHARE resource, containing a specified string of characters to be used for attribution of the Resource.

"Resource" means the language resource offered to You under the terms of this Licence. Words in the singular include the plural and vice versa.

2. The Rights Granted

2.1 The Licensor grants to You a worldwide, royalty-free, non-exclusive, licence to Use the Resource within the META-SHARE network for the duration of the META-SHARE membership

So you may, for example

- copy the Resource, or create Derivatives, or incorporate it into a Collective Work within the META-SHARE network;
- copy Derivatives, or the Resource as incorporated in any Collective Work within the META-SHARE network; and
- publish, perform or communicate the Resource and/or Derivatives and/or the Resource as incorporated in any Collective Work to anyone within the METASHARE network;

in any medium whether now known or created in the future.

2.2 However, this Licence does not allow you to:

- use the Resource in any way primarily intended for commercial advantage or payment; but exchange for other protected material without payment (whether by means of digital file-sharing or otherwise) is not to be taken to be so intended;
- impose any terms or any technological measures on the Resource or a Derivative, that alter or restrict the terms of this Licence or any rights granted under it or have the effect or intent of restricting the ability of any person to exercise those rights;
- sublicense the Resource; or
- subject the Resource to Derogatory Treatment.

2.3 You must, if you publish or distribute the Resource or any Derivative to anyone else in any way, give reasonable credit to the Original Author as follows:

! by giving the Attribution Data as well as the name of the Original Author if that has been supplied; or a pseudonym if that has been supplied instead; and/or the name of any other person if that has been supplied for attribution in place of the Original Author;

- by giving the title of the Resource if that has been supplied;
- by giving the Uniform Resource Identifier of the Resource if that has been supplied, but you need not do so if this does not refer to its copyright terms or to its licensing information;
- in the case of a Derivative, by identifying that the Resource has been used (for example, "French translation of X by Y", or "Screenplay based on original work by X"); and
- in the case of a Derivative or a Collective Work, placing that credit in the same place, and at least as prominently, as any comparable authorship credit.
- In the case of a Derivative or a Collective work to inform the Licensor about its creation

But, if what you are publishing or distributing is a Derivative or a Collective Work, you must remove any of these credits if you are asked to do so by the Licensor and if it is practicable to do so.

These are important conditions of this Licence, and if you fail to comply with them you will be in material breach of its terms.

2.4 The Original Author asserts the right to be identified as the original author of the Work, as forming part of the Resource if applicable; but only as and when required by clause 2.3 above.

2.5 You must also, if you publish or distribute the Resource or any Derivative to anyone else in any way:

- include a copy of this Licence (or its Uniform Resource Identifier) with it; and
- keep intact any copyright notices for the Resource and notices that refer to this Licence.
- Redeposit the Derivative work in the META-SHARE network

These are important conditions of this Licence, and if you fail to comply with them you will be in material breach of its terms.

2.6 Each time You publish the Resource or any Derivative to anyone else within METASHARE in any way, the Licensor offers to the recipient a licence to the Resource on the same terms and conditions as this Licence.

2.7 Any Derivative you create must be under a licence which is either one with the same terms and conditions as this Licence, or a later version of this Licence with the same Licence Elements as this Licence, or another META-SHARE licence with the same Licence Elements as this Licence (whether a licence specific to a particular jurisdiction or not).

2.8 And:

- the right to collect royalties or other fees for any commercial use of the Resource is reserved;
- any right to collect payments via a licensing body or collecting society for any commercial use of the Resource is reserved;
- the right to collect payments under the Public Lending Right scheme (or any public scheme that provides payment for public borrowing or use) is reserved;
- the right to release the Resource under different terms, or to stop distributing the Resource, is reserved; and
- all other rights not expressly granted by the Licensor are reserved.

2.9 This Licence does not affect any rights that You or anyone else may independently have under any applicable law (including fair dealing, fair use, or any other legally recognised limitation or exception to copyright infringement) to make any Use of this Resource.

2.10 This Licence does not allow You to claim any endorsement or approval by the Licensor or the Original Author of You or your use of the Resource without their express written permission.

3. Warranties and Disclaimer

3.1 The Resource is licensed by the Licensor "as is" and without any warranty of any kind, either express or implied, whether of title, of accuracy, of fitness for purpose, or otherwise.

4. Limit of Liability

4.1 Subject to any liability which may not be excluded or limited by law, the Licensor shall not be liable for, and expressly excludes, all liability for loss or damage however and whenever caused to anyone by any Use under this Licence, whether by You or by anyone else, and whether caused by any fault on the part of the Licensor or not. If liability may not be excluded by law, it is limited to actual and direct financial loss to the extent it is caused by proved gross negligence on the part of the Licensor.

5. Termination

5.1 Any breach by You of the terms of this licence (for example, by distributing the Resource without attributing as appropriate) entitles the Licensor to terminate your Licence with immediate effect and without notice to you. Persons who have received the Resource, Derivatives, or Collective Works from You under this Licence, however, will not have their licences terminated provided their use is in full compliance with this Licence or a licence granted under clause 2.6 of this Licence, i.e. if they are META-SHARE members. The waiver of database rights provided for by clause 3.1 shall continue notwithstanding any such termination.

5.2 If You are not in breach of the terms of this licence, the Licensor may not terminate your rights under it.

5.3 Unless terminated under clause 6.1, this Licence is granted to you for the duration of any rights in the Resource.

6. General

6.1 If any provision of this Licence is held to be invalid or unenforceable, that shall not affect the validity or enforceability of the remainder of the terms of this Licence.

6.2 This Licence is the entire agreement between the parties with respect to the Resource licensed here but hereby allows for additional agreements that grant more rights than this Licence. It replaces any earlier understandings, agreements or representations with respect to the Resource not specified here.

6.3 If You are in breach of the terms of this Licence (for example, by distributing the Resource without attributing as appropriate) you will not be entitled to rely on the terms of this Licence or to complain of any breach by the Licensor.

6.4 If there is any dispute as to the meaning or effect of any provision of this Licence, it must so far as possible be read and given effect in a way that is compatible with the provisions of any subsequent version of the META-SHARE Commons licence, which has the same

Licence Elements, and any interpretation of clauses and terms must take into account what is laid out in the most recent version of the META-SHARE Charter and MoU.

6.5 As far as arbitration processes have been established within META-SHARE, any dispute arising in connection with this Licence or the Resource has adhere to these processes before being filed at public justice bodies.

The limitations of suggested META-SHARE Commons licence could be noticed in many parts of this licence, so we will point out just few of them that may be important for different ACCURAT partners and their plans to use ACCURAT project results during the project and after it is finished.

Ad 2.1: “The Licensor grants to You a worldwide, royalty-free, non-exclusive, licence to Use the Resource within the META-SHARE network for the duration of the META-SHARE membership”.

In the text of the license nothing is said about the usage outside of META-SHARE. Is it allowed or not by this licence? Since this licence regulates only the usage within the membership (whose: licensor, or licensee or both?) and it is non-exclusive, we assume that the usage outside the META-SHARE membership could be regulated by some other licencing schemes.

Indeed, this has been suggested as possibility in the last column of the tabular presentation of META-SHARE Commons licences at META-NET web site (<http://goo.gl/j0zO4>).

Table 2 Presentation of META-SHARE Commons (MSC) licences

Licence name	Licence attribute	Redistribution within META-SHARE only?		Attribution required?		What purpose the original can be used for?		Derivatives allowed?		Derivatives shared under same terms?		Fee for licence?		If redistribution outside META-SHARE, under which CC licence
		yes	no	yes	no	com-mercial	non-com-mercial	yes	no	yes	no	yes	no	
MSC BY	BY	+		+		+		+			+		+	CC BY
MSC BY SA	BY SA	+		+		+		+		+			+	CC BY SA
MSC BY ND	BY ND	+		+		+			+	N/A	N/A		+	CC BY ND
MSC BY NC SA	BY NC SA	+		+			+	+		+			+	CC BY NC SA
MSC BY NC	BY NC	+		+			+	+			+		+	CC BY NC
MSC BY NC ND	BY NC ND	+		+			+		+	N/A	N/A		+	CC BY NC ND

Since ACCURAT has committed to distribute its results through META-SHARE platform, for distribution through this distribution channel and for usage by META-SHARE members, it should respect the licensing mechanism that is recommended there.

Although in META-NET D6.1.1 the META-SHARE Commons licence is quoted as only text of licence, it looks like it actually represents the basic licensing layer or bottom-line from which the licensing schemes are to be derived. Apart of META-SHARE Commons licence

combinations presented above, there are proposals for eight new schemes that are still being under consideration. These are:

- META-SHARE Non-Commercial No-redistribution licence for free (free non-commercial use and prohibited redistribution)
- META-SHARE Non-Commercial license No-redistribution licence for-a-fee (for-a-fee non-commercial use and prohibited redistribution)
- META-SHARE Non-Commercial No-redistribution No-derivatives licence for free (free non-commercial use, prohibited redistribution and prohibited derivatives)
- META-SHARE Non-Commercial license No-redistribution No-derivatives licence for-a-fee (for-a-fee non-commercial use, prohibited redistribution and prohibited derivatives)

These four will have their commercial counterparts too, thus coming to an overall number of eight. All of them will actually be editable to a certain extent, so they would serve more as licensing templates than the fixed licences. In the time of writing this report, these licensing mechanisms were not fixed yet, so ACCURAT will have to keep a close eye on their development until the end of the project and decide on appropriate licensing scheme by that time.

After careful reading it can be seen that META-SHARE Commons licence allows incorporation of “Resource” or “Derivative” into the “Collective Work”. Limitations on its distribution (for free or for fee) are more relaxed and this gives the opportunity to include fee for usage of ACCURAT resources or tools, providing that consortium agrees upon that by the end of the project. ACCURAT industrial partners, who have already expressed their intentions to exploit ACCURAT results not just for research, but also for commercial purposes, could be interested that even this possibility is open and that they could use it if they decide so.

However, ACCURAT partners are entitled to execute their right to non-exclusivity and have licencing mechanism prepared for distribution to users who are not members of META-SHARE for whatever reason, and yet want to use ACCURAT results for their own purposes. This may have different impact on licensing of language resources than to language tools collected or developed within ACCURAT project.

5 Third party licensing information

In the process of producing ACCURAT tools, primarily the ACCURAT Toolkit (Deliverable D2.6), partners had to use different third party tools and libraries in order to achieve the desired results. Table below provide information on used third party tools used for development of ACCURAT Toolkit along with information on their licensing information. This information is also relevant for recommendation of applicable licensing scheme for distribution of ACCURAT tools if they impose certain restrictions.

From the Table it can be seen that 19 tools use 20 different third party tools in 35 cases with 10 different licencing schemas. Distribution of different licensing schemas:

- 7 uses GNU GPL (v2 or later, i.e., v3) licence,
- 7 uses Apache 2.0 licence,
- 4 uses publicly available Ms-RSL,
- 1 uses CC BY v3,
- 1 uses MIT licence,
- 1 uses Wordnet licence
- 14 uses of proprietary licences, generally compatible with Apache 2.0 but with requests for special approval for commercial usage

Table 3 Third party licensing information ordered by tools

ACCURAT tool	Partner	Third party tool	Licensing schema or formulation	IPR holder(s)	Comment
DicMetric (ComMetric)	CTS	WordNet	WordNet 3.0 license (http://wordnet.princeton.edu/wordnet/license/)	WordNet 3.0 Copyright 2006 by Princeton University. All rights reserved.	Permission to use, copy, modify and distribute this software and database and its documentation for any purpose and without fee or royalty is hereby granted, provided that you agree to comply with the following copyright notice and statements, including the disclaimer, and that the same appear on ALL copies of the software, database and documentation, including modifications that you make for internal use or for distribution. WordNet 3.0 Copyright 2006 by Princeton University. All rights reserved.
		JWI (MIT Jawa WordNet Interface)	CC BY v3.0 (http://projects.csail.mit.edu/jwi/license.html)	MIT	JWI is freely available for all purposes, as long as proper acknowledgement is made.
		Stanford POS Tagger	GNU GPL v2 or later	Stanford University	The tagger is licensed under the GNU General Public License (v2 or later). Source is included. Source is included. The package includes components for command-line invocation, running as a server, and a Java API. The tagger code is ls dual licensed (in a similar manner to MySQL, etc.). Open source licensing is under the full GPL, which allows its use for research purposes, free software projects, etc. For distributors of proprietary software, commercial licensing with a ready-to-sign agreement is available. (http://nlp.stanford.edu/software/tagger.shtml)
FeaturesExtractor-Classifier	USFD	SVMlight	The program is free for scientific use. Please contact me, if you are planning to use the software for commercial purposes. The software must not be further distributed without prior permission of the author.	Thorsten Joachims, thorsten@joachims.org , Cornell University, Department of Computer Science	If you use SVMlight in your scientific work, please cite as: T. Joachims, Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning, B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, 1999. (http://www.joachims.org/publications/joachims_99a.pdf)
		Diacritics removal Unicode::UCD	All software on this site is copyright Ben K. Bullock. All rights reserved.	Ben Bullock ben.bullock@lemoda.net	Copyright © Ben Bullock 2009-2011. All rights reserved. For comments, questions, and corrections, please email Ben Bullock (ben.bullock@lemoda.net) (http://www.lemoda.net/perl/strip-diacritics/index.html)
		Transliteration tool		ILSP	Tool produced by ACCURAT partner
EMACC	RACAI	No third party IPR			

ACCURAT tool	Partner	Third party tool	Licensing schema or formulation	IPR holder(s)	Comment
PEXACC	RACAI	GIZA++	GNU GPL v2 or later	Yaser Al-Onaizan, Jan Curin, Michael Jahr, Kevin Knight, John Lafferty, Dan Melamed, David Purdy, Franz Och, Noah Smith, and David Yarowsky	"EGYPT Toolkit for Statistical Machine Translation (...) This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details. You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111-1307, USA."
		String::Similarity (Perl package)	GNU GENERAL PUBLIC LICENSE, V 2	Marc Lehmann schmorp@schmorp.de http://home.schmorp.de/	
ME-Extract	DFKI	C++ library for MaxEntropy classification	Own copyright regulation.	Tsujii Laboratory, The University of Tokyo. All rights reserved.	Redistribution and use in source and binary forms, with or without modification, are permitted for non-commercial purposes provided that the following conditions are met: 1) Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer. 2) Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
		Numerical optimization program	Own copyright regulation.	University of Chicago	This program discloses material protectable under copyright laws of the United States. Permission to copy and modify this software and its documentation is hereby granted, provided that this notice is retained thereon and on all copies or modifications.
TildeNER	Tilde	Stanford Named Entity Recognizer	GNU GPL v2 or later	Stanford University	Stanford Named Entity Recognizer is a Java implementation of a Conditional Random Field sequence model, together with well-engineered features for Named Entity Recognition. Commercial usage requires separate license http://techfinder.stanford.edu/technology_detail.php?ID%3D24628
		TreeTagger	Non-commercial usage only (as stated in the license) http://www.ims.uni-stuttgart.de/~schmid/Tagger-Licence	Helmut Schmid, Institut fuer maschinelle Sprachverarbeitung (IMS) Universitaet Stuttgart Azenbergstr. 12 D-70174 Stuttgart, Germany Helmut.Schmid@ims.uni-stuttgart.de	As the tool may not be distributed (as the license says): "The licensee has no right to give or sell the system to third parties without written permission from the licensor." we will exclude the module from the deliverable.

ACCURAT tool	Partner	Third party tool	Licensing schema or formulation	IPR holder(s)	Comment
OpenNLPWrapper	USFD	OpenNLP	Apache 2.0	Apache Software Foundation	Apache OpenNLP is an effort undergoing incubation at The Apache Software Foundation (ASF), sponsored by the Apache Incubator. (http://incubator.apache.org/opennlp/)
		Lucene	Apache 2.0	Apache Software Foundation	Apache Lucene(TM) is a high-performance, full-featured text search engine library written entirely in Java. (http://lucene.apache.org/java/docs/index.html)
		KEA	GNU GPL v3	Digital Libraries and Machine Learning Labs, Computer Science Department, The University of Waikato, Private Bag 3105, Hamilton, New Zealand	KEA is implemented in Java and is platform independent. (http://www.nzdl.org/Kea/) It will be replaced by another tool developed within ACCURAT.
NERAI	RACAI	.NET Framework 4.0	Publicly available under Microsoft Reference Source License (Ms-RSL)	Microsoft	
Tilde's wrapper system for CollTerm	Tilde	TreeTagger	Non-commercial usage only (as stated in the license) http://www.ims.uni-stuttgart.de/~schmid/Tagger-Licence	Helmut Schmid, Institut fuer maschinelle Sprachverarbeitung (IMS) Universitaet Stuttgart Azenbergstr. 12 D-70174 Stuttgart, Germany Helmut.Schmid@ims.uni-stuttgart.de	As the tool may not be distributed (as the license says): "The licensee has no right to give or sell the system to third parties without written permission from the licensor." we will exclude the module from the deliverable.
		CollTerm	Apache 2.0	Nikola Ljubešić, FFZG	The tool is developed within ACCURAT project and is published under Apache 2.0 licence.
KEA TEWrapper	USFD	OpenNLP	Apache 2.0	Apache Software Foundation	Apache OpenNLP is an effort undergoing incubation at The Apache Software Foundation (ASF), sponsored by the Apache Incubator. (http://incubator.apache.org/opennlp/)
		Lucene	Apache 2.0	Apache Software Foundation	Apache Lucene(TM) is a high-performance, full-featured text search engine library written entirely in Java. (http://lucene.apache.org/java/docs/index.html)

ACCURAT tool	Partner	Third party tool	Licensing schema or formulation	IPR holder(s)	Comment
		KEA	GNU GPL v3	Digital Libraries and Machine Learning Labs, Computer Science Department, The University of Waikato, Private Bag 3105, Hamilton, New Zealand	KEA is implemented in Java and is platform independent. (http://www.nzdl.org/Kea/) It will be replaced by another tool developed within ACCURAT.
CollTerm	FFZG	No third party IPR			
Terminology Extraction	RACAI	.NET Framework 4.0	Publicly available under Microsoft Reference Source License (Ms-RSL)	Microsoft	
Mapper	USFD	OpenNLP	Apache 2.0	Copyright © 2010 The Apache Software Foundation	Apache OpenNLP is an effort undergoing incubation at The Apache Software Foundation (ASF), sponsored by the Apache Incubator. (http://incubator.apache.org/opennlp/)
		Lucene	Apache 2.0	Copyright © 2010 The Apache Software Foundation	Apache Lucene(TM) is a high-performance, full-featured text search engine library written entirely in Java. (http://lucene.apache.org/java/docs/index.html)
		KEA	GNU GPL v3	Digital Libraries and Machine Learning Labs, Computer Science Department, The University of Waikato, Private Bag 3105, Hamilton, New Zealand	KEA is implemented in Java and is platform independent. (http://www.nzdl.org/Kea/) It will be replaced by another tool developed within ACCURAT.
NERA2	RACAI	.NET Framework 4.0	Publicly available under Microsoft Reference Source License (Ms-RSL)	Microsoft	
Terminology Aligner	RACAI	.NET Framework 4.0	Publicly available under Microsoft Reference Source License (Ms-RSL)	Microsoft	
P2GACC	LT	No third party IPR			
Translation API	CTS	Google translation API	Own copyright regulation.	Google	Google Translate API v2 is now available as a paid service only, and the number of requests your application can make per day is limited. (http://code.google.com/intl/hr/apis/language/translate/overview.html)
		Bing translation API	Own copyright regulation.	Microsoft	Bing Terms of Use (https://ssl.bing.com/webmaster/developers/tou.aspx)
DEACC	RACAI	.NET Framework 2.0	Publicly available under Microsoft Reference Source License (Ms-RSL)	Microsoft	

ACCURAT tool	Partner	Third party tool	Licensing schema or formulation	IPR holder(s)	Comment
FMC (Focused Monolingual Crawler)	ILSP	Bixo7 web mining toolkit	MIT license	© Copyright 101tec, Inc. 2008. All Rights Reserved.	The MIT License (http://www.opensource.org/licenses/mit-license.php) is a free software license originating at the Massachusetts Institute of Technology (MIT). This license requires the following copyright and permission notices to be included in all copies or substantial portions of FMC (the Software, hereafter). Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so. By respecting Bixo's license, we comply with the terms and conditions for using all the software packages on which Bixo package depends.

6 Recommended licensing schemes

After the analysis of META-SHARE licensing schemes and third party IPR for tools and libraries used in the development of ACCURAT tools, we can come up with recommendations on licensing schemes applicable to ACCURAT results.

6.1 *META-SHARE users*

Since ACCURAT has committed to distribute its results through META-SHARE platform, for distribution through this distribution channel and for usage by META-SHARE members, it should respect the licencing mechanism that is recommended there when it get stabilised. This recommendation should be applicable to both, language resources and language tools, providing that the owners of IPR on original text samples allow such licensing scheme (closed to a controlled community with membership regulation).

The preferred schema will be selected by the end of the ACCURAT project depending on the available META-SHARE licensing mechanisms.

6.2 *Non-META-SHARE users*

For users who are not members of META-SHARE, the Apache 2.0 licensing scheme is preferred for tools. The reason for this is that in seven cases of third party tool/library usage GNU GPL v2 or higher (i.e. v3) was used. In the same time proprietary licensing schemes are similar with the Apache 2.0 (with additional request for commercial use approval) and since Apache 2.0 was used also in 7 cases and since it is compatible with GNU GPL v3, our recommendation would be to use Apache 2.0 as an overall licensing scheme for all ACCURAT tools.

However, some exceptions exist, primarily with USFD FeaturesExtractorClassifier which uses SVMlight (with proprietary licence, but needs a permission by author for further distribution) and Diacritics removal script (with proprietary licence and obligation to contact the author). USFD should provide approvals by IPR owners in both cases in order to clear out the redistributionability of ACCURAT tools.

Several tools in the ACCURAT toolkit require data lemmatization and POS-tagging. Two tools from Tilde provide integration with TreeTagger for this purpose. As TreeTagger has a proprietary licence that does not allow redistribution and commercial usage, TreeTagger has been excluded from the toolkit. However guidelines for user's lemmatization and POS-tagging tool integration are provided in the Deliverable D2.6.

Regarding language resources, for non-META-SHARE users a pragmatic approach could be adopted. ACCURAT partners could make corpora collected during the ACCURAT project publicly available upon request providing that users give assurance that they have sorted our IPR issues with the original IPR owners. This position gives opportunity for interested users to fully exploit ACCURAT language resources and yet respect IPR of original text providers.

However, there will always be clear cases (e.g. South-East European Times portal that explicitly allows copying, usage and redistribution of texts from that web site providing that a fair reference is always submitted) and not so clear cases (e.g. WaC approach that crawls and processes texts from web sites producing an output that could be considered as "reading" of publicly available web site content out of which an original text should not be reconstructed). ACCURAT position is that all such language resources would be made available to users on-line if they provide assurance of sorting out IPR issues with original IPR owners.

Another practical solution could be that ACCURAT provides API access to existing resources thus limiting the quantity of textual data that could be processed in one turn. However, for this solution, we should develop API for accessing our resources what was not planned originally by the DoW.

7 Conclusion

This report on IPR issues tries to give recommendations how to treat IPR of ACCURAT results. It examines and describes existing licensing schemes that could be applicable for ACCURAT results, particularly ones proposed by META-NET, i.e., META-SHARE Commons licence, and analyses the licensing mechanisms of third party tools used in development of ACCURAT tools. At the end this reports give several recommendations how to formulate licences of ACCURAT results by the end of the project.