

**UNIVERSITY OF ZAGREB  
UNIVERSITY COMPUTING CENTRE**



**REPORT**

**Analysis of Software for Plagiarism  
Detection in Science and  
Education**

**Zagreb, September 2016**

Authors: Tamara Birkić, Draženko Celjak, Marko Cundeković, Sabina Rako

Proofreader: Mia Kožul

In Zagreb, September 9<sup>th</sup> 2016

Class: 650-03/16-421/018

Ref. No. 3801-4-421-01-16-1

Team Leader:

Assistant Director for Education and  
User Support:

Tamara Birkić, prof.

Sandra Kučina Softić, dipl.ing., v.r.



This material is available under the International Creative  
Commons License 4.0  
*Attribution-NonCommercial-NoDerivs.*

<http://creativecommons.org/licenses/by-nc-nd/4.0/deed.hr>



# CONTENT

<b>1. INTRODUCTION .....</b>	<b>5</b>
<b>2. PLAGIARISM DETECTION SOFTWARE .....</b>	<b>7</b>
<b>3. CRITERIA AND ANALYSED SOFTWARE .....</b>	<b>8</b>
3.1. API and plug-in .....	9
3.2. Service location .....	10
3.3. Database scope and the possibility to include personal content .....	10
3.4. Support.....	12
3.5. Distribution .....	12
3.6. Costs of using the software and licence.....	13
3.7. Authentication and user roles .....	15
3.8. Software testing and quality assessment .....	16
<b>4. CONSOLIDATED TESTING RESULTS.....</b>	<b>19</b>
<b>5. CONCLUSION.....</b>	<b>20</b>
<b>6. APPENDIX – API OVERVIEW.....</b>	<b>22</b>
6.1. PlagScan .....	22
6.2. Turnitin .....	23
6.3. Unplag.....	23
6.3. Urkund.....	24
<b>7. REFERENCES.....</b>	<b>25</b>

# 1. INTRODUCTION

Plagiarism, i.e. ethics in education in general, has been a frequent topic of discussion in recent years. In 2015, aware of the trends in education, European Council established a Pan-European Platform for Ethics, Transparency and Integrity in Education (ETINED)<sup>1</sup>. One of its missions is to protect, develop and support the academic integrity, especially the fight against plagiarism. Academic integrity among higher education institutions is highly important due to the increasing number of students, as well as the growing competition among universities.

The European Commission has also recognized the importance of this topic and in the period from 2010 to 2013 it carried out the IPPHEAE project (Impact of Policies for Plagiarism in Higher Education Across Europe) with a goal to research policies and systems put in place to ensure academic integrity and prevent plagiarism in higher education (Glendinning, 2015). As a part of this project the Academic Integrity Maturity Model (AIMM) was also created.

The AIMM model identifies these criteria (Glendinning, 2014) on which the national evaluations of 19 countries of the EU were based on<sup>2</sup>:

- transparency in academic integrity and quality assurance
- fair, effective and consistent policies for handling plagiarism and academic dishonesty
- standardisation of sanctions for plagiarism and academic dishonesty
- use of digital tools and language repositories
- preventative strategies and measures
- communication about policies and procedures
- knowledge and understanding about academic integrity
- training provision for students and teachers
- research and innovation in academic integrity.

From the criteria listed above it is evident that the plagiarism issue is a complex one and it needs to be reviewed from multiple perspectives (organizational and technical). A comprehensive system that can prevent plagiarism can be established by coordinating and involving these identified elements.

Plagiarism detection software is an important element of a systematic plagiarism detection. Plagiarism detection software has many advantages, such as the possibility to review a large volume of papers stored in repositories in a short period of time, to do similarity checks as well as to create reports that can be used as certificates of originality.

---

<sup>1</sup> <http://www.coe.int/en/web/ethics-transparency-integrity-in-education>

<sup>2</sup> Croatia was not included in the IPPHEAE project so there is no available data which could help determine the state of the Croatian higher education system in relation to other European countries.

In Croatia, there are several examples of universities systematically addressing the issue of plagiarism, but the impression is that, in order to resolve the issue, it should be approached with an even better strategy and more systematically.

With this document, University Computing Centre (SRCE) wants to encourage discussion on the issue of plagiarism in higher education with a focus on plagiarism detection software, as well as present and share gathered information with colleagues and the general public. Information that SRCE gathered can be a valuable input to higher education institutions that are choosing a certain plagiarism detection software, but it can also stimulate discussion about the necessity of a solution at the university or even at the national level.

Due to the fact that SRCE maintains national e-infrastructure systems DABAR – Digital Academic Archives and Repositories, Merlin – University e-Learning Platform and HRČAK – Portal of Scientific Journals in Republic of Croatia, results of this analysis may serve as SRCE's recommendation to the higher education institutions how to detect and prevent plagiarism using computer software.

## 2. PLAGIARISM DETECTION SOFTWARE

A dictionary definition of plagiarism is „an act or instance of using or closely imitating the language and thoughts of another author without authorization and the representation of that author's work as one's own, as by not crediting the original author“.<sup>3</sup> As it is already stated in the introduction, one of the ways of preventing plagiarism is by using plagiarism detection software.

Why use plagiarism detection software? What can it offer to the academic community?

The importance of plagiarism software detection is evident in many examples of good practice in the world. Croatia has a small number of institutions that use this type of software.

By analysing a paper, the software offers the user information on how much of the content is identical, i.e., copied from other sources. However, it is important to point out that there is no defined boundary that will separate original papers from plagiarised ones. It is up to the user to carefully and with understanding interpret the results obtained from the software analysis. Also, the results of a software analysis cannot be interpreted in the same way in all science fields. Many common technical terms or mathematical formulas describing them cannot be considered plagiarism, although they appear in various papers. For this reason, teachers are expected to review the results of the analysis and to give a definite conclusion on whether something is plagiarised or not.

Universities that had implemented the plagiarism detection software noticed that the use of software also contributes to the raise of awareness about ethics among students and that students pay more attention while paraphrasing, referencing or citing the work of other authors (Stappenbelt and Rowles, 2009).

Methods of using plagiarism detection software vary. In some cases, the originality of the paper can be validated only by a faculty member, while in others students check their own papers before handing them in.

It is important to point out that plagiarism detection software has certain limitations and therefore can still only recognize plagiarised text, but not plagiarised multimedia (pictures, video, etc.). Software mostly compares strings from the paper being analysed with other papers stored in its database, which means that a certain tool is as good as the database it uses (the quality of the database depends on the quality and amount of papers stored in it). Also, plagiarism detection software cannot detect plagiarism in different languages (translations).

---

<sup>3</sup> <http://www.dictionary.com> (report in Croatian language referes to the definition given by <http://hjp.znanje.hr>)

## 3. CRITERIA AND ANALYSED SOFTWARE

The goal of the analysis conducted by SRCE was to gather information on the technology used by plagiarism detection software and the possibilities it offers, as well as to create recommendations for the academic and research community in Croatia. The first step was creating criteria, which were followed with identifying the most commonly used software in academic environment and analysing it. The software was chosen according to the possibility of usage on the national or institutional level (universities or faculties) and considering relevant sources in Croatian language (HRČAK, DABAR), as well as relevant information systems that would be able to apply the software (Merlin – University e-Learning Platform).

Defined criteria:

- available application programming interface (API) and plug-in in order to integrate the software into DABAR and Merlin systems
- service location (locally on institutions' or SRCE servers or online on software manufacturers' servers)
- database scope and the possibility to include personal content
- support
- distribution (usage of software in the world, Europe and neighbouring countries – Slovenia, Bosnia and Herzegovina, Serbia)
- costs of using the system and licence
- authentication using an AAI@EduHr user account and methods of authentication (multiple roles within the system).

After determining the criteria, all the plagiarism detection software available on the market were reviewed. The following software was selected and analysed, due to its high usage in Europe:

- PlagScan
- StrikePlagiarism
- Turnitin
- Unplag
- Urkund.

Software Strike Plagiarism was not considered further after the initial analysis indicated a lack of key functionalities.<sup>4</sup>

---

<sup>4</sup> Only one document can be reviewed at a time, there are difficulties uploading large files and PDF files are not supported.

During the analysis of selected software, contacted manufacturers provided requested information about each software according to the defined criteria via online presentations. Also, each selected software and its possibilities were tested.

Criteria for software testing were:

- identification of cited parts of the text
- supported formats and restrictions
- intuitive interface (subjective evaluation and ratings from 1 to 10, where 1 is the lowest and 10 the highest grade).

### 3.1. API AND PLUG-IN

The objective of this part of the analysis was to determine to what extent can the plagiarism detection software be integrated with national and other systems (e.g. DABAR and Merlin) via programming interface (API) and Moodle and Drupal plug-in. Support, the possibility of parameterization and availability of examples and documentation were also very important factors. An important criterion for Moodle plug-in was the possibility to enable or disable software usage to certain institutions (Merlin system is used by many universities and faculties in Croatia).

<b>PLAGSCAN</b>	API is available, as well as the documentation and implementation examples for Java, PHP and .NET programming languages. (web site: <a href="http://www.plagscan.com/api-guide">http://www.plagscan.com/api-guide</a> ). API results, in a form of an analysed document with highlighted parts that need to be checked, as well as statistics (percentage of matching...), are also available. Moodle plug-in is publicly available. Drupal plug-in is not available, but the manufacturer is willing to develop it, if necessary.
<b>TURNITIN</b>	Software does not have an API, but it is compliant with the <a href="#">Learning Tools Interoperability (LTI)</a> standard which enables integration into e-learning systems. Moodle plug-in ( <a href="#">Moodle Direct</a> ) is publicly available.
<b>UNPLAG</b>	API is available, as well as the documentation ( <a href="https://unplag.com/api/doc/">https://unplag.com/api/doc/</a> ). Parameter <i>similarity sources</i> , which omits sources with less than 5% similarity, is very useful. Moodle plug-in is publicly available. While testing it in the Merlin system, it could not be successfully installed.
<b>URKUND</b>	API is available, but the documentation is given only upon request. Moodle plug-in is available and it was successfully installed in the Merlin system. This is the only plug-in with the possibility of restricting the usage of software to certain institution in the Merlin system used by several institutions.

### 3.2. SERVICE LOCATION

This criterion was used to compare the available options of accommodating the software and the database: **online** on software manufacturer's server or **locally** on institution's or SRCE's server. The advantage of an online service is that the software user does not have to plan for or maintain additional computing resources, while the advantage of a local service is that the user has a full control over the data.<sup>5</sup>

<b>PLAGSCAN</b>	online Possibility of local server installation ( <i>PlagScan in a Box</i> package) which is charged a one-time \$ 4,000 per installation and \$ 99 a month for maintenance. <sup>6</sup>
<b>TURNITIN</b>	online No possibility of local server installation.
<b>UNPLAG</b>	online For documents stored on an institutional and/or associate institutions level ( <i>My Library</i> ) the software can be installed on a local server (costs range from \$ 10,000 to \$ 20,000). <sup>7</sup>
<b>URKUND</b>	online No possibility of local server installation.

### 3.3. DATABASE SCOPE AND THE POSSIBILITY TO INCLUDE PERSONAL CONTENT

Database scope refers to the sources the software uses to detect plagiarism. To identify plagiarism, it is necessary for the original work to be included in the database the software uses, i.e., successfulness of the software at detecting plagiarism depends on the scope and quality of the database. It is preferable that the database includes or can include relevant Croatian Internet sources (DABAR repository, HRČAK). It is extremely important that the institution can add papers from their own database to the software's database. In that situation, it is also important to check the conditions under which the content is added.

<b>PLAGSCAN</b>	Four sources: <ol style="list-style-type: none"> <li>1. the Internet (Microsoft Bing search engine and selected academic web sites)</li> <li>2. personal database (all institutional and users' documents)</li> <li>3. publications and journals (around 21,000 scientific journals)</li> <li>4. PlagScan database (papers stored in the software's database with the prior approval of the author or the person who analysed the paper).</li> </ol>
-----------------	--

<sup>5</sup> Software manufacturers define rights on the papers which are submitted to the plagiarism check.

<sup>6</sup> Price was quoted during a meeting held in March 2016, but the information is also available via software's web site: <http://www.plagscan.com/in-a-box-local-server>.

<sup>7</sup> Price was quoted during a meeting held in January 2016.

	<p>The web interface offers the option of storing papers in the software's database.</p> <p>Institutions can include their own databases and repositories. In the event that the access to the content is protected, PlagScan can access them via user interface or available API.</p> <p>There is a possibility of including papers from HRČAK and DABAR.</p>
<b>TURNITIN</b>	<p>Three sources:</p> <ol style="list-style-type: none"> <li>1. the Internet (commercial search engine for academic web sites)</li> <li>2. scientific articles (this software includes the largest number of journals compared to all the other analysed software)</li> <li>3. users' papers included in the database (users are not given the choice to delete the paper from Turnitin database).</li> </ol> <p>The software includes more than 60 billion web sites, 600 million student papers and 154 million books and journals.</p> <p>Analysed papers are indefinitely included in their database and cannot be deleted. There is a possibility to limit the visibility of the document by choosing the <i>Invisible to others</i> option (during the analysis it will be pointed out that there are similarities with a certain source, but the source or the author(s) will not be named).</p> <p>HRČAK is included as the source and upon request, so can DABAR.</p>
<b>UNPLAG</b>	<p>Two sources:</p> <ol style="list-style-type: none"> <li>1. the Internet (Microsoft Bin and Yahoo search engines, as well as certain academic web sites)</li> <li>2. personal database (the software database can include institutional and/or associate institutions' papers (<i>My Library</i>)).</li> </ol> <p>An individual analysis is also possible (comparing two documents). While uploading papers, there is a possibility of setting documents' access rights (global-everyone, institution, teacher, student). While deleting the document from the database, the global index is also deleted (if it was available).</p> <p>Papers from HRČAK and DABAR can also be included (upon request for permission to include repositories).</p>
<b>URKUND</b>	<p>Three sources:</p> <ol style="list-style-type: none"> <li>1. the Internet (the software uses its own crawler and index, i.e., it does not use search engines)</li> <li>2. papers published in the Urkund database (since 2014, 11.5 million papers in total)</li> <li>3. personal database (all institutional and user's documents).</li> </ol> <p>Papers can be excluded from the Urkund database upon request.</p> <p>HRČAK and DABAR are included as sources.</p>

### 3.4. SUPPORT

While working in the system, institutional users find the availability of ongoing support very important. Apart from the user support, development support and customisation support are also important as they ensure stability and long-term sustainability of the system. This kind of support does not include support to end users (students, teachers).

<b>PLAGSCAN</b>	Standard support (included in the price of the software) includes: <ol style="list-style-type: none"> <li>1. online support (e-mail, webinars), phone (9am to 5pm) – expected response time is 24 hours</li> <li>2. if needed, meetings at the customer's location.</li> </ol> Additional customisations of the system are charged extra, unless there is a large number of interested parties for the same customisation.
<b>TURNITIN</b>	Standard support (included in the price of the software) includes: <ol style="list-style-type: none"> <li>1. regular support via phone, e-mail or online form.</li> </ol>
<b>UNPLAG</b>	Standard support (included in the price of the software) includes: <ol style="list-style-type: none"> <li>1. regular support via phone, e-mail or YouTube channel</li> <li>2. key account manager – institution's personal support available via mobile phone.</li> </ol> Additional customisations of the system are charged extra (the price depends on the scope of the work).
<b>URKUND</b>	Standard support (included in the price of the software) includes: <ol style="list-style-type: none"> <li>1. regular support via phone, e-mail or Skype – expected response time is 24 hours.</li> </ol>

### 3.5. DISTRIBUTION

Plagiarism detection software is used in different parts of the world so one of the criteria was the size of the user community, especially in Europe. Data presented in the table were obtained in the first half of 2016.

<b>PLAGSCAN</b>	According to the software manufacturer, there are no institutional users in Croatia. The largest customers in Europe are universities in Germany (1200 institutions – 800 schools and 400 universities), Austria, Ukraine, Spain, Cyprus and Switzerland. University of Maribor is the only software user from the neighbouring countries and it has been using the software for two years.
<b>TURNITIN</b>	According to the software manufacturer, the software is used by the University of Rijeka, University of Osijek, VERN' University of Applied Sciences and the Zagreb School of Economics and Management while several faculties of the University of Zagreb have started negotiations. More than 10,000 institutions around the world (135 countries) use this software which makes it the most commonly used software.

<b>UNPLAG</b>	<p>According to the software manufacturer, there are no institutional users in Croatia although negotiations with several faculties of the University of Zagreb have started.</p> <p>The largest customers in Europe are universities in Germany, Italy, Spain and the UK, and the world base includes around 150 customers.</p>
<b>URKUND</b>	<p>According to the software manufacturer, there are no institutional users in Croatia.</p> <p>The largest customers in Europe are universities in Italy, Portugal, Germany, Sweden, Norway, Austria, and in the world North America.</p> <p>This software increased the number of end-users by 81,000 in the last five months, while the total number of users is several million.</p> <p>Negotiations are under way with several universities in Serbia.</p>

### 3.6. COSTS OF USING THE SOFTWARE AND LICENCE

When deciding on the software, payment model by which the price is formed (the number of users, number of pages, number of documents, number of characters per page, etc.) is very important, as well as the possibility of software usage and costs formation individually per institution. Tables below present software comparison and prices.

<b>PLAGSCAN</b>	<p>Two payment models:</p> <ol style="list-style-type: none"> <li>1. number of students (unlimited number of analysis)</li> <li>2. number of analysed pages/words (one page contains 275 words).</li> </ol> <p>Licence duration: one year.</p> <p>The price is publicly available at software's web site and has been clearly defined.</p> <p>Analysis of scientific journals is included in the price if the higher education institution publishing the journal is the software user.</p>
<b>TURNITIN</b>	<p>Two payment models:</p> <ol style="list-style-type: none"> <li>1. number of students (unlimited number of analysis)</li> <li>2. number of analysed pages/words.</li> </ol> <p>Licence duration: one year.</p> <p>To analyse scientific journals a different system should be used, called iThenticate, which is charged individually.</p>
<b>UNPLAG</b>	<p>Two payment models:</p> <ol style="list-style-type: none"> <li>1. number of students (unlimited number of analysis)</li> <li>2. number of analysed pages/words (one page contains 275 words).</li> </ol> <p>Licence duration: one year.</p> <p>To analyse scientific journals, it is necessary to inform the manufacturer on the amount of analysis per year to determine whether the analysis can be included at no additional cost or for an additional fee.</p>
<b>URKUND</b>	<p>Three payment models:</p> <ol style="list-style-type: none"> <li>1. number of students (unlimited number of analysis)</li> <li>2. number of analysed pages/words</li> </ol>

	<p>3. number of documents (50% of the costs is payed up front, while the rest in instalments or at the end of the year). Licence duration: one year. Analysis of scientific journals is included in the price if the higher education institution publishing the journal is the software user. For other journals that wish to use the software independently, licence costs 800 € plus 2 € for each analysed document.</p>
--	---

Prices according to payment methods (*quoted April 20<sup>th</sup> 2016*). The lowest prices are highlighted green.

Software	Number of students (yearly licence)			
	150,000	50,000	6,000	2,000
PLAGSCAN	0.85 €	0.92 €	1.00 €	1.00 €
TURNITIN*	0.80 €	0.80 €	1.50 €	1.50 €
UNPLAG	2.00 €	2.10 €	2.25 €	/
URKUND	0.49 €	0.68 €	1.00 €	1.20 €

\* Quoted prices are valid if negotiations are done with the Ministry of Science, Education and Sports of the Republic of Croatia.

Software	Number of pages (yearly licence)	
	100 million words / 18,939 €	200 million words / 36,582 €
PLAGSCAN	100 million words / 18,939 €	200 million words / 36,582 €
TURNITIN	/	/
UNPLAG	35,000 pages / 0.05 €	70,000 / 0.04 €
URKUND	/	/

Software	Number of documents (yearly licence)	
	35,000	70,000
PLAGSCAN	/	/
TURNITIN	/	/
UNPLAG	/	/
URKUND	1 € /per document/max.length 400,000 characters	0.75 € /per document/max. length 400,000 characters

### 3.7. AUTHENTICATION AND USER ROLES

It is important that teachers and students can use the plagiarism detection software by logging in with their existing identities within the AAI@EduHr system. By using the existing infrastructure of the AAI@EduHr system there is no need to create and maintain a potentially large number of user accounts within the software.

During the analysis, the number of different types of user roles that exist in the software were taken into account (system administrator, institutional administrator, mentor, student...), as well as the possibility to set role and user quota.

<b>PLAGSCAN</b>	<p>AAI@EduHr e-ID is enabled (Shibboleth / SAML / Active Directory are supported).</p> <p>Additionally, teacher can generate a key which the student uses to upload papers using the software's web site, therefore there is no need to sing into software.</p> <p>Roles:</p> <ul style="list-style-type: none"> <li>• administrator</li> <li>• subadministrator (institutional administrator)</li> <li>• teacher</li> <li>• student.</li> </ul> <p>Administrators can create user accounts individually or in large groups. Administrators can also set a quota for certain users – how many words can a user check for plagiarism. The system uses points, where 1 PlagPoint is 100 words.</p>
<b>TURNITIN</b>	<p>AAI@EduHr e-ID is enabled (Shibboleth is supported).</p> <p>Roles:</p> <ul style="list-style-type: none"> <li>• administrator</li> <li>• teacher</li> <li>• student.</li> </ul>
<b>UNPLAG</b>	<p>There is no support for Shibboleth/SAML, the software requires local user accounts.</p> <p>Roles:</p> <ul style="list-style-type: none"> <li>• administrator</li> <li>• teacher</li> <li>• student.</li> </ul>
<b>URKUND</b>	<p>AAI@EduHr e-ID is enabled (Shibboleth is supported).</p> <p>Roles:</p> <ul style="list-style-type: none"> <li>• administrator</li> <li>• teacher</li> <li>• student.</li> </ul>

## 3.8. SOFTWARE TESTING AND QUALITY ASSESSMENT

Each software was tested using a sample of 10 papers from the following categories:

- articles published in journals
- conference papers
- doctoral dissertations
- M.A. theses
- student papers.

Two papers from each category were selected, one written in the period from 2005 to 2006 and the other from 2015 to 2016.

The amount of recognized copied content was not significantly different depending on the software the paper was checked with, so it is safe to assume that the analysed software are similar. The processing speed of each document depends on the amount of text it contains. Documents were processed in a similar amount of time by each analysed software.

Tables below present the results of software testing:

### a) PLAGSCAN

<b>Quote recognition</b>	- cited text is recognized according to the similarities within quotation marks - quote recognition option can be disabled
<b>Limitations</b>	- analysis results available in PDF or DOCX format - downloaded Word document marks copied parts (potential plagiarisms) via comments
<b>Interface intuitiveness</b>	9
<b>Notes</b>	- when certain sources, which were initially recognized as plagiarized, are disabled, the software calculates a new percentage - intervals of similarity percentage can be adjusted and given a colour attribute (e.g. 20 – 30% green, 30 – 60% yellow etc.) - whitelist enables excluding certain web sites from the analysis - possibility of excluding pre-defined text from the analysis (e.g. "Name and surname")

## b) TURNITIN

<b>Quote recognition</b>	- quote recognition option can be disabled and certain references/bibliography excluded from the analysis using smart filters
<b>Limitations</b>	- does not recognize special characters (č, ć, š, đ, dž, ž) so the words containing them are excluded from the report - student can send only one document from his/her computer, Dropbox or Google Drive for analysis
<b>Interface intuitiveness</b>	7,5
<b>Notes</b>	- possibility of adding voice comments to students, grading using sections, indicating errors using drag-and-drop method with already predefined errors (in the form of marks) that are frequent in student papers - teachers can send quick feedback to students (teachers define standard replies/comments) - university statistics is recorded (number of students, teachers, reports, number of documents grouped by percentage of plagiarism, possibility to export statistics into Excel) - possibility to adjust the content of the report (availability of the report to the student, minimal threshold in percentage or words that the system will recognize as plagiarism)

## c) UNPLAG

<b>Quote recognition</b>	- quotes are recognized according to the quoting style (MLA, APA...) - testing gave the impression that each bracket is considered a quote - possibility of including quoted text
<b>Limitations</b>	- there were none noticed
<b>Interface intuitiveness</b>	7,5
<b>Notes</b>	- the software detects plagiarism even when certain letters in the copied text are replaced by an alternative font (e.g. if the English a is replaced by a Russian one)

d) URKUND

<b>Quote recognition</b>	- does not recognize quotes, but there is a possibility to enable text recognition within quotation marks and brackets
<b>Limitations</b>	- analysis of one PDF document was unsuccessful (technical issues) - only documents with more than 450 characters can be analysed
<b>Interface intuitiveness</b>	8
<b>Notes</b>	- the software recognizes and marks parts of the document that were copied from papers written in Serbian language - the interface does not offer a possibility of choosing a single source of verification (e.g. Internet, another document or database) - when the software finds more sources with the same text, only the most common sources are listed, while the rest, sources with a smaller percentage of similarity, can be found in the <i>Sources not used</i> box - when certain sources are excluded, reduced percentage is not saved, instead the initial percentage generated by the software is set

## 4. CONSOLIDATED TESTING RESULTS

The table below presents consolidated testing results which enable a quick overview of each system according to the defined criteria.

Legend:

-  (not fulfilled or below average)
-   (fulfilled)
-    (additional options available).

Criterion	PlagScan	Turnitin	Unplag	Urkund
API AND PLUG-IN				
PlagScan stands out due to its well documented and quality API. Urkund Moodle plug-in is the only one with a possibility of restricting the usage of software to certain institution in the Merlin system used by several institutions.				
SERVICE LOCATION				
PlagScan and Unplag have an additional possibility of being installed on a local server.				
DATABASE SCOPE AND THE POSSIBILITY TO INCLUDE PERSONAL CONTENT				
PlagScan and Urkund allow the user full control over the documents being analysed. Turnitin's advantage is a large number of journals included in the database, while its disadvantage is that the papers are indefinitely stored in the software's databased and cannot be deleted.				
SUPPORT				
Each software includes user support. From the contacts we made, we would like to point out we had good experience with representatives of PlagScan and Unplag software.				
DISTRIBUTION				
Turnitin is used by several higher education institutions in Croatia.				
COSTS OF USING THE SOFTWARE AND LICENCE				
Urkund is the cheapest and Unplag the most expensive.				
AUTHENTICATION AND USER ROLES				
PlagScan allows institutional administration, while Unplag does not support integration with the AAI@EduHr system.				
SOFTWARE TESTING AND QUALITY ASSESSMENT	9/10	7,5/10	7,5/10	8/10

## 5. CONCLUSION

The final selection of plagiarism detection software largely depends on the needs of an individual institution. In this document, we have tried to give an overview of information we find important, but also to show results and observations gathered during the testing of each particular tool.

The possibility to integrate a tool with an existing information systems via API or plug-ins is an important feature of plagiarism detection software. This analysis was particularly concerned with the possibility of integration with DABAR, HRČAK and Merlin (Moodle) systems. These systems are frequently used in the academic community and are maintained by SRCE. All the analysed plagiarism detection software support some kind of integration (API, LTI or Moodle plug-in). A closer look at the available documentation discovered minor differences among software – the impression is that the PlagScan system has a more detailed documentation that can speed up and facilitate the integration. The PlagScan documentation also provides improved query parameters operability. In addition, when considering the possibility of integration with the Merlin (Moodle) system, a scenario was defined in which the plagiarism detection system would be available only to certain institutions that use Merlin, not to the whole system. The only software that enables this scenario is Urkund.

This analysis is mainly based on the available documentation and presentations given by software representatives. For the definite functionality verification of API and Moodle plug-in, software should be installed and tested in the system. Usage scenarios should be determined as well.

One of the important issues is how to keep control over documents analysed by plagiarism detection software. Each analysed software offers the possibility of sending documents online via their infrastructure, while PlagScan and Unplag also offer the possibility of installing the software locally, but at extra cost. During the plagiarism check, all the systems browse Internet resources and locate documents from databases such as HRČAK and DABAR. All software representatives are open for suggestions and requests to include new index sources. All tools have the possibility to add or enable personal content (i.e. papers from individual institutions) in the software's database, but Turnitin pointed out that student papers cannot be deleted from their database.

Considering the distribution of software usage at the moment this analysis was created, only Turnitin is used at some of the higher education institutions in Croatia.

The costs of using the system vary and depend on the number of students that will use the software. The larger the number of students, the cheaper the costs. Some software manufacturers are open to negotiations over the price. All licences are valid for a year, starting from a specified date (not necessarily the beginning of a calendar or academic year).

Because SAML/AAI@EduHr authentication is used by all higher education institutions in Croatia, information whether software supports this kind of authentication or not was very useful. The number and rights of user roles (administrator, teacher students) were also

important. PlagScan system has a well developed administrator interface with a number of useful options and adjustable set of parameters, while Unplag does not support authentication via AAI@EduHr e-ID.

It is important to point out that tools do not provide a result which defines whether some document is plagiarised or not. They are used to find similarities with other documents, while it is up to the teacher or the person in charge to review the results of the analysis with understanding.

Software testing conducted as a part of this analysis gave us a better insight into the possibilities of each tool, but it is important to mention that software is regularly upgraded and the observations described in this documents should be considered in the context of the time in which the testing took place. This document analysed four commercial software, but one should take into account that due to the rapid development of technology new commercial and open source software will appear.

Communication with all software representatives was great, except with the Turnitin representative who allowed testing only after prolonged negotiations.

**After the analysis, as serious candidates for use in science and higher education system, we recommend PlagScan or Urkund.** To reach a final decision, an additional testing should be conducted that would include the user community (teachers, students) and representatives of institutions. At the same time, it is necessary to adopt policies that would ensure organizational preconditions for the application of plagiarism detection software.

## 6. APPENDIX – API OVERVIEW

### 6.1. PLAGSCAN

<https://www.plagscan.com/api-guide>

Documentation: Online, sample code for Java, PHP, .NET.

Security requirements:

- SSL
- restricted IP addresses (one or several IP addresses).

Return formats:

- statistics only (plagiarism level, word count)
- document content and result links
- XML with links to found sources
- Docx document with marked plagiarisms
- HTML document with marked plagiarisms
- HTML report
- PDF version of the HTML document and report.

Results parameters:

- analysed document's PlagScan ID, user's ID, word count, date, analysis status (paused, processing, completed, in queue)
- plagiarism level, file name
- first 85 characters of content.

Configuration parameters:

- language (English, German, Spanish)
- setting the „yellow“ and „red“ plagiarism border (in percentages)
- email notification (never, always, only if „red“ plagiarism level, i.e. high level of plagiarism)
- generate Docx documents (generate and email, generate only, do not generate)
- autostart plagiarism checks (yes / no)
- check against the Internet for plagiarism (yes / no)
- compare to (no one / my documents / my institution / general database)
- analysis sensitivity (low / medium / high)
- automatically remove data after (one week, four weeks, six months, never).

## 6.2. TURNITIN

Supports Learning Tools Interoperability (LTI):

[https://guides.turnitin.com/03\\_Integrations/Learning\\_Tools\\_Interoperability\\_\(LTI\)](https://guides.turnitin.com/03_Integrations/Learning_Tools_Interoperability_(LTI))

Note: Uses (*Learning Tools Interoperability*) instead of standard API

Documentation: Online

## 6.3. UNPLAG

<https://unplag.com/api/doc/>

Documentation: Online PDF (accessible with a registered user account)

Security requirements:

- SSL
- restricted IP addresses (one or several IP addresses).

Return formats:

- JSON, XML, MsgPack

Parameters:

- Store document
  - file format
  - file
  - file name
- Delete document
  - file's ID
- Start analysis
  - ID of the file being analysed
  - ID of the analysis files (Docs-vs-Doc)
  - Exclude citations (yes / no)
  - Exclude references (yes / no)
- Obtain file info (ID, Word count, Name, Format, Page number)
- Obtain analysis info (ID, Price, Type, number of compared documents, Date, Status, Progress)
- Obtain results info (Date, Similarity, Number of sources, Number of citations, Number of references, URL link)
- Generate PDF report
  - file's ID
  - language
- Get report link
- Toggle citations and references
- Track progress

Support for *Learning tools Interoperability* (LTI):

<https://www.imslobal.org/activity/learning-tools-interoperability>

## 6.3. URKUND

Documentation: Per request

Security requirements:

- SSL

Return formats:

- JSON, XML

Parameters:

- Store (Id, Date and time, File name)
- Status (Stored, Rejected, Approved)
- DocumentInfo (document info)
  - Document's Id
  - Date
  - Document link
- ReportInfo (results info if the analysis was successful)
  - Report ID
  - Report link
  - Plagiarism level
  - Number of text similarity
  - Number of sources
  - Document error warnings
- ReceiverInfo
  - Email
  - Name and surname

## 7. REFERENCES

Glendinning, I. (2015). *Promoting Maturity in Policies for Plagiarism across Europe and beyond*. Presented at: 7th Prague Forum: Towards a Pan-European Platform on Ethics, Transparency and Integrity in Education, Prague.

Glendinning, I. (2014). *Assessing maturity of institutional policies for underpinning academic Integrity*. International Integrity and Plagiarism Conference. Held from 16<sup>th</sup> to 18<sup>th</sup> June 2014 in The Sage Gateshead, UK. Available at:

<https://core.ac.uk/download/files/169/30620175.pdf> (2.7.2016)

Pan-European Platform for Ethics, Transparency and Integrity in Education (ETINED)

Available at: <http://www.coe.int/en/web/ethics-transparency-integrity-in-education>.

Stappenbelt, B. i Rowles C. (2009). *The effectiveness of plagiarism detection software as a learning tool in academic writing education*. Presented at: 4th Asia Pacific Conference on Educational Integrity (4APCEI), Wollongong. Available at:

<http://ro.uow.edu.au/apcei/09/papers/29/> (26.7.2016)



(Analysis of Software for Plagiarism Detection in Science and Education.docx)