

# **ВРЕМЕННИ ПРАВИЛА**

## **ЗА ЕЛЕКТРОННА ПРОВЕРКА НА ОРИГИНАЛНОСТТА**

### **НА ДИПЛОМНИ РАБОТИ (МАГИСТЪРСКИ ТЕЗИ)**

#### **РАЗДЕЛ I. ОБЩИ ПОЛОЖЕНИЯ**

**Чл. 1.** (1) Настоящите правила уреждат реда и условията за извършване на електронна проверка на оригиналността на дипломни работи (магистърски тези) на студентите от Университета за национално и световно стопанство (УНСС) по учебния план на всяка специалност.

(2) Целта на проверката е да се гарантира академичната почтеност и да се установи степента на оригиналност на дипломните работи (магистърски тези).

(3) Проверката е задължителна за всички студенти, завършващи образователно-квалификационна степен „бакалавър“ и „магистър“, които подлежат на защита на дипломна работа (магистърска теза) в УНСС.

#### **РАЗДЕЛ II. РЕД ЗА ИЗВЪРШВАНЕ НА ПРОВЕРКАТА**

**Чл. 2.** (1) Електронната проверка на оригиналността се извършва **от студента**, който е автор на дипломната работа (магистърската теза) след одобрение на научния ръководител на завършена разработка.

(2) За целта студентът подава **електронно заявление за проверка**, чрез установената електронна система, определена от УНСС на следния линк: <https://forms.office.com/e/auj8iGPsR4?origin=lprLink>.

(2) Към заявлението се прилага завършената дипломната работа (магистърската теза) в Word и/или PDF формат.

(3) Катедрите и научните ръководители е необходимо да предвидят период от минимум 10 работни дни за да бъде направена проверката за оригиналност на дипломната работа (магистърската теза)

(4) След извършване на проверката, оторизирани лица от персонала на библиотеката на УНСС изпращат доклад, генериран от софтуера до студента и до научния ръководител чрез служебната електронна поща.

(5) След направената проверка за оригиналност и тълкуване на индикаторите (Приложение 1), научният ръководител взема решение дали да допусне студента до защита на дипломна работа (магистърска теза) на текущата дипломна сесия.

#### **РАЗДЕЛ III. ПРОВЕЖДАНЕ И ПОКАЗАТЕЛИ НА ПРОВЕРКАТА**

**Чл. 3.** (1) Проверка на оригиналността се извършва чрез утвърдена от УНСС електронна система за анализ на текстови сходства и цитиране – Strike Plagiarism, а в Приложение 1 е представено тълкуване на индикаторите.

(2) Системата генерира доклад, който съдържа следните основни показатели:

1. Коефициент на сходство 1 (КС1) – процент на текстово сходство от минимум 5 думи с други източници;

2. Коефициент на сходство 2 (КС2) – процент на сходство от минимум 25 думи с други източници;

3. Коефициент на цитати (КЦ) – процент на цитирани текстове спрямо общия обем на дипломната работа (магистърска теза).

#### **РАЗДЕЛ IV. КРИТЕРИИ ЗА ДОПУСТИМОСТ ДО ЗАЩИТА**

**Чл. 4.** (1) Всяка катедра определя праговите стойности по критериите за допустимост до защита на дипломни работи (магистърски тези), според Приложение 1.

(2) В случай, че стойностите по някой от показателите надвишават посочените допустими граници, спрямо изискванията на катедрата, научният ръководител взема решение дали да допусне студента до защита на дипломна работа (магистърска теза) на текущата дипломна сесия, въз основа на доклада и тълкуване на индикаторите.

#### **РАЗДЕЛ V. ДОКУМЕНТИРАНЕ И СЪХРАНЕНИЕ НА РЕЗУЛТАТИТЕ**

**Чл. 5.** (1) Катедрата съхранява докладите като част от официалната документация за защита на дипломната работа (магистърска теза).

#### **РАЗДЕЛ VI. ЗАКЛЮЧИТЕЛНИ РАЗПОРЕДБИ**

**Чл. 6.** Компетентен орган за извършване на електронна проверка на оригиналността на дипломните работи (магистърски тези) чрез специализиран софтуер е сектор „Информационни ресурси“ към дирекция „Университетска библиотека“ на УНСС. Секторът поддържа софтуера и базата данни, необходими за извършване на проверките.

**Настоящите правила са приети от Академичния съвет на Университета за национално и световно стопанство № \_\_\_\_ от \_\_\_\_\_ 20\_\_ г.**

## Проверка за оригиналност и тълкуване на индикаторите

### Препоръчителни прагови стойности на коефициентите за сходство:

- ✓ Коефициент на сходство 1 (КС1) – 20%-30%;
- ✓ Коефициент на сходство 2 (КС2) – 10%-15%;
- ✓ Коефициент на цитати (КЦ) – 10%-20%.

КС1 и КС2 са помощни индикатори, които насочват вниманието към евентуални заемки, но не се тълкуват самостоятелно като доказателство за сходство.

Дори висок коефициент на сходство (особено КС1) не бива автоматично да се приравнява на плагиатство без допълнителен анализ. Винаги е необходимо компетентно лице (преподавател или проверяващ) да прегледа доклада в детайли – да разгледа източниците на съвпаденията и характерът на съвпадналите фрагменти. Например, ако съвпаденията идват от общоизвестни фрази, клишетата или терминология, наличието им не означава непременно плагиатство.

КС1 обхваща и по-кратки фрази, затова е по-чувствителен към често срещани изрази. В резултат на това е възможно дори оригинални трудове да имат осезаем КС1, поради използването на стандартни фрази или термини. Анализите показват, че на английски език разпространени фрази като „*както беше споменато по-рано*“ или „*може да се заключи, че*“ могат да съставляват до ~10% от работата, създавайки „шум“ в КС1. Сходна ситуация се наблюдава и при текстове, наситени със специфична терминология – те по-често имат по-висок КС1, въпреки че съдържанието им може изцяло да е оригинално. Поради това изолирано висок КС1 (над 50%) изисква внимателна проверка, но не е сам по себе си доказателство за плагиатство. При висок КС1 и нисък КС2, често причината е именно множество дребни съвпадения (по 5-10 думи) – случайни или обусловени от темата фрази, които не представляват умишлено преписване.

КС2 е по-стриктен показател – той регистрира само дълги буквални съвпадения (праг от 25 думи). В практиката на StrikePlagiarism се приема, че висока стойност на КС2 е сериозен сигнал за сходство. Причината е, че статистически малко вероятно *идентични фрази от 25 думи* да се появят в два различни документа, освен ако единият не е заимствал от другия. Както се отбелязва в ръководствата, наличието на поне един такъв дълъг идентичен откъс е „*достоверен признак за заемка*“. Затова, ако КС2 надхвърля примерно 50%, проверяващият задължително трябва внимателно да разгледа съвпаденията и от какви източници идват.

КЦ измерва дела на текста, подаден в кавички, спрямо целия документ. Изчислението става като системата преброява думите (или символите) във всички открити цитирани фрагменти и ги съотнася към общия брой думи в документа, получавайки процент. Например, ако от 10 000 думи в един ръкопис 2 000 думи са в кавички (цитати), то коефициентът на цитиране ще бъде 20%.

*Важно:* StrikePlagiarism отчита само цитати, които са коректно форматиращи, т.е. пасажки, започващи и завършващи с кавички. За да бъде един фрагмент включен в коефициента на цитиране, трябва ясно да личи, че е цитиран. Основният критерий е употребата на кавички:

- Целият заемен пасаж трябва да бъде ограден в кавички „...“ (или съответния формат на кавички за съответния език), за да го отчете софтуерът като цитат.

- Вътрешно цитиране (цитат в рамките на друг цитат) също би било разпознато, стига форматването да е правилно (напр. единични кавички вътре в двойни, ако се следва стандартен стил).

- Бележки под линия, които съдържат препратки към източници, се считат за част от цитатите, ако съдържат текст в кавички или стандартни референтни фрази.

Високото съдържание на цитати само по себе си не е плагиатство, ако цитатите са правилно оформени и по възможност подкрепени с референции. Например, ако докладът показва  $KC1=40\%$  сходство, но в същото време  $KI=30\%$ , това подсказва, че до 3/4 от установените съвпадения може да са цитирани абзаци. В такава ситуация проверяващият трябва да провери дали тези цитати са надлежно придружени от източници (бележки, библиография) и дали са употребени коректно.

StrikePlagiarism изрично препоръчва внимателно анализиране на съдържанието – да се провери „дали цитатите са маркирани и дали произлизат от документи, посочени в списъка с референции“.

Трябва да се има предвид, че цитираните пасажки (текст, поставен в кавички и обозначен като цитат) могат да бъдат изключени от изчислението на коефициентите. StrikePlagiarism.com позволява на проверяващия да изключи открития цитиран текст от  $KC1$  и  $KC2$ , тъй като такъв текст не се брои за плагиатство. Това означава, че ако даден открит фрагмент е правилно цитиран и се приеме като легитимен цитат, може да се изключи от калкулацията – тогава процентите  $KC1/KC2$  автоматично спадат.

Ако цитиранията са легитимни, високият процент сходство не бива автоматично да се счита за нарушение. От друга страна,  $KI$  насочва вниманието и към случаите на прекомерна употреба на чужд текст. Макар да не е „плагиатство“ в строг смисъл, необичайно висок коефициент на цитиране може да означава, че документът разчита твърде много на чужди думи. Академичните стандарти изискват балансирано използване на източници – цитати, но и значителен собствен принос.

В доклада от StrikePlagiarism е необходимо да се разгледа списък с най-дълги фрагменти и да прецени дали съвпаденията представляват случайни повторения/обща фрази или действително взаимствани параграфи. Например, ако в „Списък на най-дългите фрагменти“ фигурират само кратки откъси, разпръснати из текста, те може да се сметат за случайни съвпадения. Но ако се откроява един източник с голям процент и дълги пасажки, това изисква внимание – вероятно става дума за копиран текст, който следва да бъде обозначен като сходство.

*Пример:* Ако в проверяван документ от 10 000 думи системата намери общо 2 000 думи, които принадлежат на фрагменти (мин. 5 думи) от други източници, то  $KC1 = 20\%$ . Ако от тях 300 думи са в един или няколко дълги пасажа по  $\geq 25$  думи, то  $KC2 = 3\%$ . В този случай докладът би показал  $KC1=20\%$  и  $KC2=3\%$ .

Модулът за AI-детекция на StrikePlagiarism.com посочва вероятността за използване на изкуствен интелект за отделните фрагменти от документа и ги групира в 5 диапазона. Този коефициент показва каква част от текста би могла да бъде създадена с помощта на изкуствен интелект в рамките на предварително зададената от потребителя прагова стойност. 97% означава, че 97% от текста има коефициент на вероятност за съдържание, генерирано от изкуствен интелект, по-висок от 78% (0,78 - предварително определен коефициент на вероятност за изкуствен интелект). 0 - 20% 21 - 40% Човешки стил (без следи от ИИ) 41 - 60% Смесен стил (частично използване на ИИ) 61 - 80% 81 - 100% Стил с ИИ (висока вероятност за използване на ИИ).