Útmutató a Turnitin Similarity szövegegyezés-vizsgálat eredményének értelmezéséhez

A program az adatbázisába bekerülő dokumentumokkal, internetes (jelenlegi és régebbi, már nem elérhető), illetve Crossref tartalmakkal veti össze a vizsgálandó szöveget. A Turnitin a más forrásokkal egyező szavak számának és a vizsgált szöveg terjedelmének (szavak száma) arányából egy százalékban kifejezett értékkel adja meg a vizsgálat eredményeként a hasonlósági számot (%), és megjelöli a talált szövegrészeket, forrásokat.

Szövegegyezés-vizsgálatra feltölthető dokumentumok adatai:

- Az egyes fájlok mérete nem haladhatja meg a 100 MB-ot
- A fájl legalább 20 szót kell tartalmazzon
- A dokumentum maximális terjedelme 800 oldal lehet
- A támogatott fájltípusok közé tartoznak: Microsoft Word, Excel, PowerPoint, PostScript, pdf (nem írásvédett,), HTML, HWP, rtf, OpenOffice (odt), WordPerfect és egyszerű szöveg (txt).
- **megjegyzés:** a Word-ből „Nyomtatás PdF-be" funkcióval készített fájlok esetében a szöveg nagy része képpé alakul, így a szövegegyezés-vizsgálat vagy nem is fut le, vagy csak a szövegként megmaradó részeket vizsgálja. (ellenőrzés: kijelölhető a szöveg az egérrel azokon a részeken, ahol nincs találat?)

Fontos tudni, hogy a százalékos hasonlósági szám alapján nem lehet kimondani, hogy bizonyos érték felett egyértelműen plágiumról van szó, vagy sem. A dokumentumról készült jelentést minden esetben értelmezni kell, amelyhez az alábbiak adhatnak támpontot.

Legelőször ajánlatos átnézni a teljes szöveget, az alábbiakra figyelve:

1. Amennyiben a vizsgálandó dokumentum korábbi változata is fel volt töltve szövegegyezés-vizsgálatra, **a korábbi változatot ki lehet kapcsolni a vizsgálatból (Exclude source).** Indokolt esetben, a korábbi változat töröltetése az adatbázisból az adminisztrátoroktól kérhető a plagium@uni-mate.hu címen a feltöltés azonosítójának (oid) megadásával.

2. **A szöveg gyors áttekintése** következik, amelynek során már adott esetben már el is dönthető, hogy egyértelműen plagizált részek vannak benne, és a dolgozat így nem fogadható el. Ilyen esetben nem szükséges a vizsgálatot folytatni. Ha nem állapítható meg egyértelműen a plagizálás ténye, akkor tovább lehet lépni.

3. **Az irodalomjegyzék áttekintése és kikapcsolása a vizsgálatból (Settings/Exclusion Filters: Exclude bibliography)**

    a. Ha az irodalomjegyzék összefüggő, egyetlen színnel kiemelt egyezést mutat, azt jelentheti, hogy az irodalomjegyzéket egyben kimásolták valahonnan.

    b. Ha az irodalomjegyzékben egy-egy tétel nincs kijelölve semmilyen színnel, az arra utalhat, hogy a hibás, nem létezik, esetleg mesterséges intelligencia (pl. ChatGPT) hozta létre (ezek sokszor valós helyekhez hasonlítanak). Az ilyeneket érdemes mindenképp leellenőrizni.

    Magyar nyelvű dokumentumokban nem működik ez a szűrő.

4. Az **idézetek** esetén ellenőrizni kell a szerzőség jelzésének pontosságát, illetve az idézőjelek helyes használatát (pl. nincs rejtett, fehér színű idézőjel a szövegben) majd ezután kikapcsolhatók ezek is az elemzésből **(Settings/Exclusion Filters: Quoted text). Magyar nyelvű dokumentumokban a szűrő hibásan működhet a „ …" és "…" jelölés közötti különbség miatt, vagy hiányos jelölés miatt, úgyhogy bekapcsoláskor figyelni kell a % szám változását, és ellenőrizni a szövegben, ha túl nagyot ugrik ez az érték a szűrő bekapcsolása által.**

5. **Hivatkozott szövegrészek** esetén ellenőrizni kell a szerzőség jelzésének pontosságát **(Settings/Exclusion Filters: Quoted text, Cited text).**

6. **A rövidebb, pár szavas találatok kizárhatók a vizsgálatból** a szavak számának megadásával **(Settings/Exclusion Filters: Exclude small matches)**. Rövid mondatrészek egyezése esetén ugyanis előfordulhat, hogy általános kifejezésekről van szó, ezért ezek a pár szavas egyezések nem jelentenek plágiumot. **Figyelem:** Ha ugyanabból a forrásból több egyező rövid mondat, mondatrész is kijelölésre kerül, az azt jelezheti, hogy kimásolták a szövegrészt, majd itt-ott módosítottak rajta, kibővítették azt.

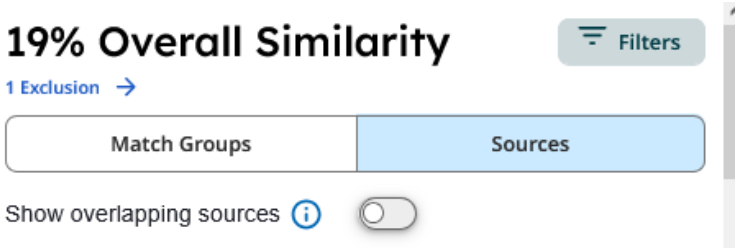
A szűrőket érdemes egyesével kapcsolni, nem egyszerre, és figyelni a hasonlósági szám (%) változását. Ha irreálisan nagy a változás, elképzelhető, hogy a szűrő valamilyen okból hibásan működik.

7. **Ezután a jelentés részletes vizsgálata következik, amely a dolgozat tartalmának értékelésével egyidejűleg történhet.** Hosszabb, egybefüggő egyező szövegrészek (néhány szó különbséggel) találata mindenképpen gyanús. Ilyen esetben meg kell vizsgálni, hogy az adott forrás fel van-e tüntetve az irodalomjegyzékben, továbbá azt, hogy annak szerzői között szerepel-e a vizsgált dokumentum szerzője.

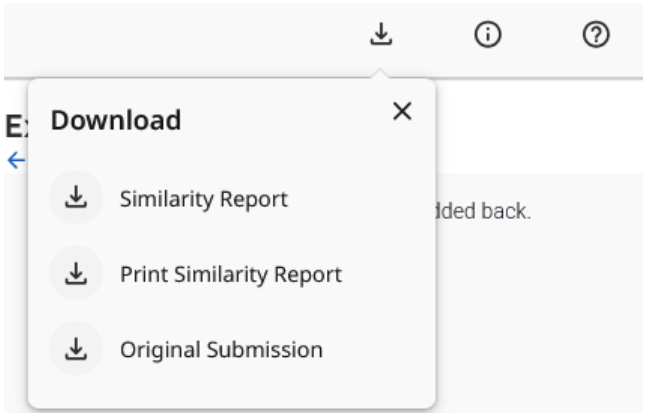
8. Ritkább esetekben előfordulhat, hogy hosszú szövegrészekben sok forrásból, teljesen egyező mondatok szerepelnek. (Más-más színnel jelölt teljes egyezés) Ez azt jelentheti, hogy mondatonként különféle forrásokból épül fel a szöveg.

9. A **piros zászló ikonnal** megjelölt részek csalási kísérletre utalhatnak. Az algoritmusok például olyan szövegrészekre hívják fel a figyelmet, ahol fehér háttéren fehér színű karakterek vannak. Például egy fehér színű idézőjelek közé zárt plagizált szövegrész az idézetek kizárásával kikerülhet a találatok közül. A hosszabb, fehér színű szöveg megemeli a szavak számát, szemmel nem látható, de a százalékos eredményt jelentősen módosíthatja. Leggyakoribb azonban, amikor ábrák, táblázatok, képletek esetén felülírás vagy más karakterek beszúrása történik.

10. **A vizsgálatból kizárt szövegek és források listája (Exclusions)** ellenőrizhető, az elemei vissza is helyezhetők szükség szerint.



11. **A vizsgálat végeredménye a nem releváns találatok kiszűrése után** pdf formátumban tölthető a jobb felső sarokban található **(Download Similarity Report) linkre kattintva.**
Megj. a Neptun-Turnitin hibás integrációja miatt Byte.byte típusú fájlok csak átnevezés után nyithatók meg. A kiterjesztés .pdf kell legyen.



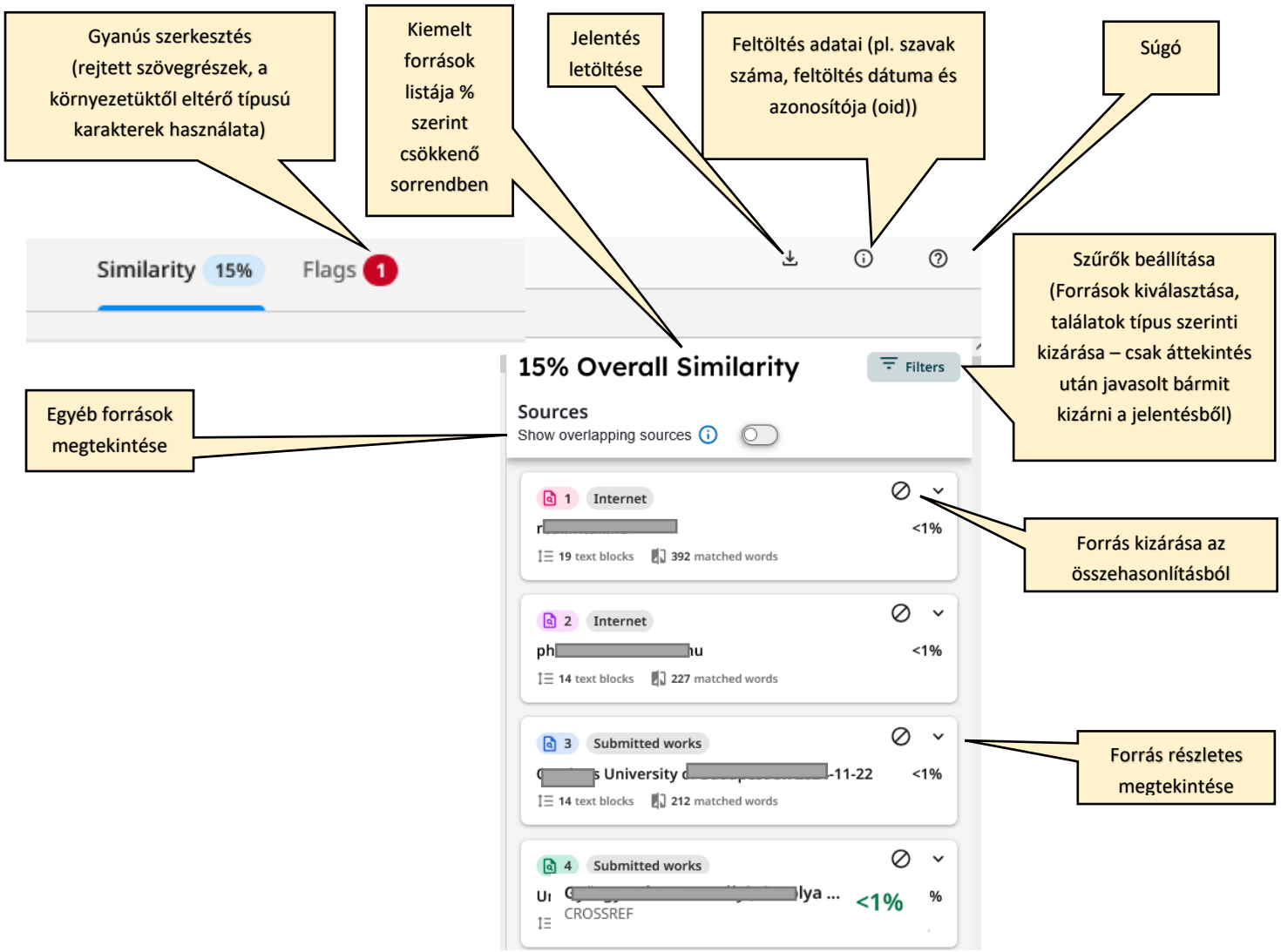
Tartalmazza a dokumentum adatait, az egyező (kiszínezett, megszámozott) szövegrészeket, a Kiemelt forrásokat, az egyezőség mértékét, illetve a kizárt források listáját is.

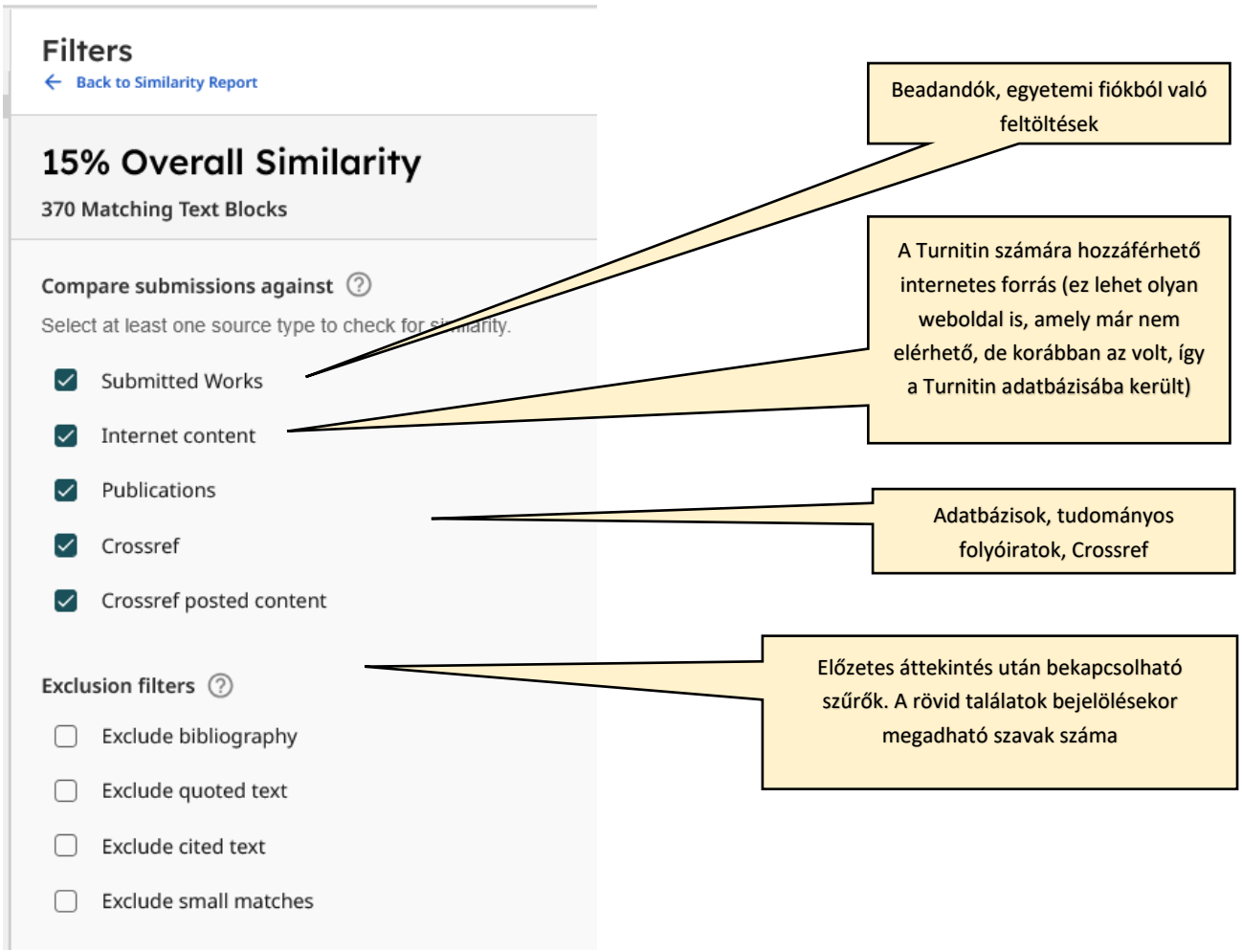
Az jelentés kezdőoldalán a **Kiemelt források** láthatóak, amelyek az összes forrás közül az egyező szavak száma, a forrás frissessége, relevancia szerint komplex algoritmusok segítségével kerülnek kiválasztásra. Érdemes egy pillantást vetni az **Összes forrásra** is (Show overlapping sources).

A Turnitin **idegen nyelvű forrásokkal** is összeveti a dokumentumot, de tudni kell, hogy az idegen nyelvről fordított szövegrészeket nem tudja jól kezelni, főleg nem a magyar nyelvű dokumentumokban.

A **Turnitin forrásai** is korlátozottak. Ha van olyan weboldal, amelyet szeretne, hogy a Turnitin hozzáadjon az adatbázisához, kérjük, töltse ki az űrlapot, vagy vegye igénybe adminisztrátoraink segítségét (plagium@uni-mate.hu).

A **mesterséges intelligencia** használatával szerkesztett vagy megalkotott szövegrészek kimutatása (AI writing detection) funkció nem áll rendelkezésünkre. Tudni kell, hogy ezek használata jelentősen megváltoztatja a szövegegyezés-vizsgálat eredményét. A magyar nyelvű szövegek esetében MI alkalmazására utalhat, ha hosszabb szövegrészben semmilyen szövegegyezést nem mutat ki az elemzés, ha nem koherens, a hivatkozások nem pontosak, az irodalomjegyzék egyes tételei nem valósak, az egész szövegben furcsa hibák, nem odavaló kifejezések tűnnek fel.





Kérdés esetén munkatársainkat az elearning@uni-mate.hu és a plagium@uni-mate.hu címen kereshetik.

Készítette: Egyetemi Könyvtár és Levéltár, 2025.04.08.