(Plágium típusai: 1.) szó szerinti másolás, 2.) tartalommásolás (parafrazálás), 3.) illetve külön esetként kezelendő a fordított (I.) és AI generált szövegek másolása.)

Indításkor megjelenik egy „Quick start” ablak, rövid szövegekkel és képekkel a főbb funkciókról, (kikapcsolható). Minden felületnek biztosítania kell az egyszerű, intuitív használatot, illetve akadálymentesítettnek kell lennie (pl. feliratozott gombok, színtévesztőknek is könnyen azonosítható grafikai elemek, vakok által használt felolvasó programok támogatása, stb). A „Súgó” menünek (benne support elérhetőséggel) könnyen megtalálhatónak kell lennie a használat bármilyen szakaszában. (Felhő alapú szolgáltatásként szükséges regisztráció és bejelentkezés funkció, de jelen dokumentum a helyi gépre telepített verziót taglalja, amelynek a licencét biztosíthatja pl. termékkulcs.)

A kezdő kezelőfelületen behivatkozhatjuk a vizsgált dokumentum útvonalát (bemásolás, ill. tallózás), vagy behúzhatjuk másik mappából (drag-and-drop).
Ha helyi adatbázist is itt hivatkozhatjuk be (bejelentkezés pl. tokennel).
Itt módosíthatunk a keresés típusán és scope-ján (szó szerinti keresés gyorsabb a tartalmi jelentésnél, vagy a helyi adatbankban az internetes keresésnél), illetve több nyelv támogatása esetén . A keresés gombra kattintva megkezdődik a szöveg előfeldolgozása, beállítás szerinti keresések, majd a találatok begyűjtése.

A folyamat befejzésekor megjelenik a dokumentumunk képe, zöld szövegháttérrel az eredetiként érzékelt, pirossal a lehivatkozás hibás (szó szerinti idézés), sárgával a parafizáltként érzékelt szövegrészek. A színezett dokumentumunk alatt listaként megjelenik a hibára okot adó anyagok elérési információ (név, cím, link/útvonal), illetve kereszthivatkozással összekötve a saját szövegünk belüli részekkel.
Az eredményt mutató felület felső részében jelenjenek meg a százalékos adatok, illetve ábrán is érzékelhető legyen az eredeti és a különböző problémás részek közötti arány.
Alul még szükséges még egy „mentés” és „megosztás” (email, intézményi adminisztrációs rendszer), illetve „visszatérés” gomb. Mentéskor (pdf) lássa el időpecséttel.

A beállítás felületen adhatjuk meg a küszöbértéket (alapérték: 3), azaz hány egymás utáni beazonosított szó számít már idézésnek, illetve vizsgálja-e a funkcionális szavakat (és, ha; and, or, stb – alap esetben nem vizsgálja). Konfigurálhatjuk, mi számít azonos jelentésű szónak (mekkora legyen a szóvektorok távolsága), példákkal prezentálva (kutya-eb vagy kutya-négylábú) – ez „advanced” beállítás, nagyon használhatatlan eredményeket okozhat a rossz beállítás, érdemes jogosultséghoz kötni. Szintén itt lehet beállítani, mi számítson kulcsszónak (mi legyen az inverz gyakoriság logaritmikus értéke ), és mi az a mennyiség és „szövegterület”, ami már parafrazálás-nak számítson (szintén jogusultsághoz kötött).

Egy általános tájékoztató felületen a programhoz kapcsolódó alap információkon (szoftver célja, felhasználása területe, álatalános tájékozatók) túl egy lista, milyen adatbázisokban keres, illetve a feedback-et és részletesebb hibajelzést is itt lehet adni, error logok beimportálásával).

Szöveg előfeldolgozása:
a.) word tokenization trigram kialakításához: központozás eltávolítása, lower case, szegmentálás. b.) tartalom leképezés (Inverse Document Frequency) c.) kulcsszavak szűrése AI generált template modellel és vektorizálása (jelentésük számszerűsítése, word2vec). A jelentés szerinti vizsgálat csökkenti a fordítások miatti false negatív eredményeket.

Vizsgálat:

1.) Ha több (küszöbértéknél több) egymás utáni trigram ugyanarra a forrásra utal, az eredeti szövegbeli szakasz elején idézőjelnek, végén idézőjelnek és lehivatkozásnak kell lennie.

2.) 2.) parafrazálás esetén, a kulcsszavak jelentésein végez mennyiségi vizsgálatot (Naive Bayes), a nullával való szorzás elkerülésével (Smoothing).

Adattárolás: nem tárol adatot, a megadott adatbázisokhoz read-only hozzáférése lehet.

Keresés scope-ja:

1.) intézmény saját adatbázisa (tantárgy, tanár, téma szerint más diákok leadott anyagai között).
2.) Interneten elérhető anyagok (wikipédia, google books, ghostwriting.services, kutatói portálok), többi intézmény elérhető adatbázisa
3.) netes keresőmotorok (Babylon Search, Bing, stb.)

AI detection:
A különböző mesterséges intelligencia által generált szövegek felismerésének kidolgozása és vázlatos bemutatása jelen dokumentumnak nem volt célja, viszont mint tovább fejlesztendő terület érdemes elgondolkodni egy már meglévő AI decetor api-val való kommunikációt, illetve megvizsgálni az ezeket is kijátszani képes további oldalak működését.

Felhasznált irodalom

(I.) Kiss András Károly: A plágiumkereső szoftverek kiskapui
http://www.irisro.org/pedagogia2013januar/0407KissAndrasKaroly.pdf
(II.) Brian Yu: CS50's Introduction to AI with Python - Language
https://cs50.harvard.edu/ai/2020/weeks/6/
(III.) Patrick Shyu („TechLEad”): Using ChatGPT with YOUR OWN Data. (LangChain OpenAI API) https://www.youtube.com/watch?v=9AXP7tCI9PI
→ suggested for test version: LlamaIndex
https://gpt-index.readthedocs.io/en/latest/index.html
(IV.) Andy Stapleton: AI Detection Bypass: Uncovering the Only Method That Works! I Tried Them All!
https://www.youtube.com/watch?v=PI9NqITSCac