

EGÉSZSÉGÜGYI ÁLHÍRKERESŐ ELJÁRÁSOK INFORMATIKAI MEGOLDÁSAI

COMPUTATIONAL TOOLS FOR HEALTH-RELATED FAKE NEWS DETECTION

Csendes Tibor¹, Berend Gábor², Németh T. Enikő³

¹az MTA doktora, Szegedi Tudományegyetem Informatikai Intézet, Szeged
Pannon Egyetem Műszaki Informatikai Kar, Veszprém
csendes@inf.szte.hu

²PhD, Szegedi Tudományegyetem Informatikai Intézet, Szeged
berendg@inf.szte.hu

³az MTA levelező tagja, Szegedi Tudományegyetem Általános Nyelvészeti Tanszék, Szeged
nemethen@hung.u-szeged.hu

ÖSSZEFOGLALÁS

Az MTA Tudomány a Magyar Nyelvért Nemzeti Program *Álhírek, áltudományos nézetek nyelvészeti azonosítása* című projektje keretében eljárásokat fejlesztünk egészségügyi álhírek automatikus felismerésére. A jelen cikkkel kettős célunk van. Egyrészt röviden bemutatjuk az álhírkereső eljárások fejlesztése terén elért első eredményeinket, azokon belül is az informatikai megoldásainkat. Másrészt szeretnénk a már használható számítógépes eszközeink példáival az ilyen munkák iránt érdeklődőket megnyerni együttműködőknek.

ABSTRACT

In the framework of the Hungarian Academy of Sciences' Science for Hungarian Language National Programme project entitled *Linguistic Identification of Fake News and Pseudoscientific Views*, we are developing computational methods for the automatic detection of health-related fake news. The aim of this paper is twofold. On the one hand, we briefly present our first results in the development of fake news detection methods, including our computational tools. On the other hand, we would like to attract colleagues interested in such work as collaborators by providing examples of our already working computational tools.

Kulcsszavak: automatikus álhírfelismerés, álhír, egészségügy, pragmatika, reguláris kifejezés

Keywords: automatic fake news recognition, fake news, health care, pragmatics, regular expression

ELŐZMÉNYEK

A Szegei Tudományegyetem (SZTE) Általános Nyelvészeti Tanszéke és a hozzá kötődő Pragmatika Hallgatói Kutatóműhely évek óta foglalkozott a manipuláció és az álhírek nyelvészeti eszközökkel felismerhető jeleinek, startégiáinak a keresésével. 2022 nyarán az SZTE Informatikai Intézetében Gáspár Tamás mesterszakos hallgató elkészített egy olyan okostelefonos alkalmazást, amely az álhírek nyelvészeti tulajdonságainak ellenőrzésére volt képes. Ugyanazon év őszén az MTA Tudomány a Magyar Nyelvért Nemzeti Programja keretében az MTA–SZTE–DE Elméleti Nyelvészeti és Informatikai Kutatócsoport kapott támogatást négy évre az *Álhírek, áltudományos nézetek nyelvészeti azonosítása* című projektje megvalósítására.

2023 novemberében zártuk az első évet, az informatikai eszközeink megjelenítése, külalakja kialakult, lehet őket tesztelni. Ez megfelel az ún. „zöld banán” fejlesztési elvnek, illetve a felhasználói élmény (user experience, UX) korai kialakításának. Az érdemi algoritmusok finomítása, továbbfejlesztése valószínűleg kitölti majd a projekt teljes idejét, de a most elérhető eljárásaink is jók és használhatók már.

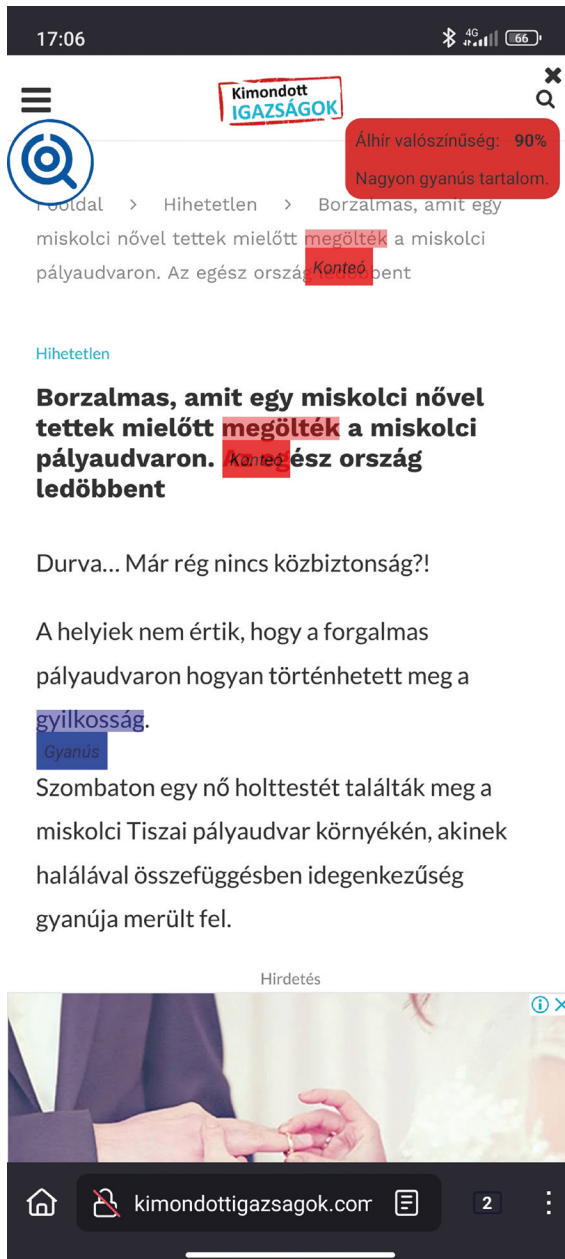
ELSŐ EREDMÉNYEK

Okostelefonos alkalmazás

Továbbfejlesztettük a fentebb említett működőképes okostelefonos alkalmazást, amely képes a nyelvészcsoporthoz által összeállított jellemzők automatikus felismerésére. Az alkalmazás a háttérben futva, a képernyőn megjelenő tetszőleges szövegen végzi a detektálást, megjelöli a gyanús részeket, rövid magyarázatot ad az okra, és összegzi a teljes látható szövegről való benyomását az álhírgyanú valószínűségének megadásával. Az alkalmazás egy *online* adatbázis segítségével dolgozik, amelyet a nyelvészek egy egyszerű kezelőfelületen tudnak bővíteni vagy javítani. Az összegzett értékelés egyelőre *ad hoc* jellegű, nincs visszamérve, illetve a valós helyzethez igazítva. Az *1. ábra* segít elképzelni az alkalmazás eredményét, amelyről egy bemutató videó is elérhető az Álhírdetektor oldalról (URL1), vagy közvetlenül a YouTube-ról is (URL2). A későbbiekben tervezzük ennek összekötését a mesterségesintelligencia-alapú álhírfelismerő programmal. Az alkalmazás maga is letölthető az Álhírdetektor oldalról (URL1), Android 8 és későbbi rendszereken működik.

Mesterségesintelligencia-alapú felismerő motor

Az álhírfelismerésre egy mesterséges neuronhálón alapuló módszert is fejlesztettünk ötezer egészségügyi álhírral és ugyanennyi korrekt egészségügyi szöveggel való betanítással. Ez egy magyar nyelvi modell segítségével dolgozik, a 110 millió



1. ábra. Az álhírfelismerő okostelefonos alkalmazás működés közben

Megjegyzés: A program csak a kijelzőn megjelenő szöveget használja. Ezen megjelöli a gyanúra okot adó jeleket, és a jobb felső sarokban színnel és százalékos értékkel is jelzi az álhírnyanú mértékét.

paraméteres HuBERT (Nemeskey, 2020) mesterséges neuronháló finomhangoláson átesett változata hozza meg a döntést. A HuBERT-modell használatán túl a kompaktabb HuBERTUSz-modellcsalád (Ficsor–Berend, 2023) használatával is kísérleteztünk, amellyel hasonlóan jó eredményekre sikerült jutnunk. Álhírdetektáló modelljeink az ún. transzformer neurálháló-architektúrát alkalmazó nyelvi modellre támaszkodnak. Ennek az a lényege, hogy a program a szöveget alkalmas rövid jelsorozatokra bontja, és az érdemi elemzés ezeken történik. A transzformer modellek az utóbbi időben sok természetesnyelv-feldolgozási (natural language processing, NLP) probléma megoldásával bizonyították sikeres alkalmazhatóságukat, mivel ez a fajta architektúra lehetővé teszi a szövegekben rejlő belső összefüggések, kontextusok és nyelvi jelenségek feltárását, amelyek rendkívül hasznosak lehetnek változatos, a szövegek feldolgozásával kapcsolatos feladatok elvégzésében, így az álhírek azonosításában is.

Tesztelésünk szerint a modell kb. 90%-os pontossággal dolgozik, jelenleg egybekezdésnyi szövegen hatékony. Ez a pontosság az átlagos emberi megítélésnél kicsit jobb az általunk vizsgált korpuszon. A pontos kiértékelést az 1. táblázat mutatja a tanítási és tesztelési adathalmazunk alapján. A táblázatban szereplő F1 mutató a pontosság és a fedés harmonikus közepe. A pontosság itt a rendszer által helyesen azonosított álhírek aránya az összes, a rendszer által álhírként azonosíthatóhoz képest. A fedés pedig a helyesen azonosított álhírek aránya az összes álhírhez képest.

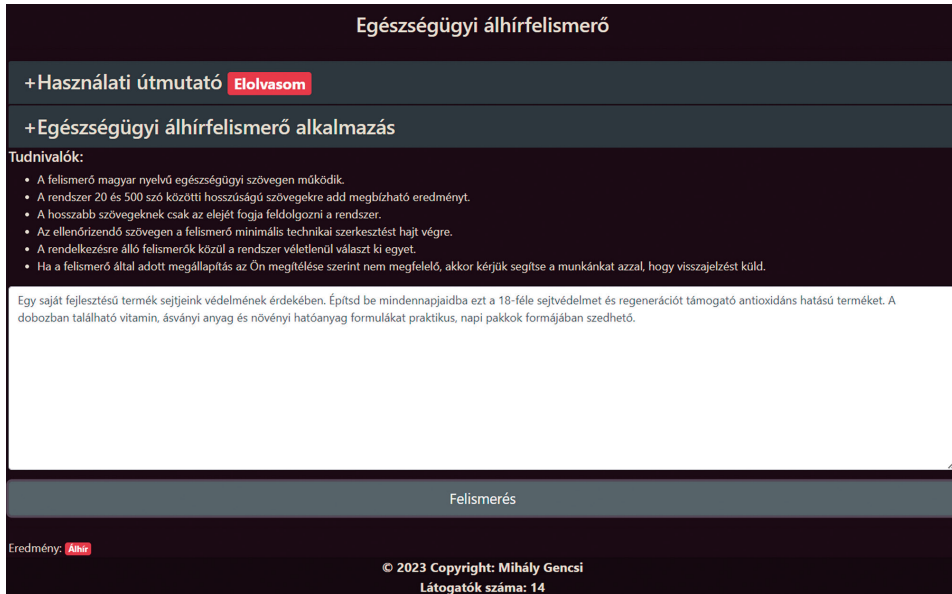
1. táblázat. A kifejlesztett mesterséges neuronháló teljesítménye az F1 mutató szerint címekre, illetve tartalomra használva, négy különböző nagyságrendű kapacitással rendelkező magyar nyelvi modell szerint

	HuBERTUSz-tiny	HuBERTUSz-small	HuBERTUSz-medium	HuBERT
Paraméterek száma	4,5 millió	29,5 millió	42 millió	110 millió
Címekre	0,9399	0,9423	0,9423	0,9549
Tartalomra	0,9196	0,9272	0,9322	0,9422

A modellhez tartozik egy visszajelzést támogató üzenetküldő funkció is. Az ezzel szerzett tapasztalattal tudjuk javítani a program eredményességét. A program általában jól működik, akár a végleges megoldásunk is lehet.

A projektünk hátralévő két évében leginkább ennek az eszköznek a továbbfejlesztését szeretnénk végezni. A többi informatikai eszköz vagy készen van, vagy épp a mesterséges intelligencián alapuló fejlesztésekre vár. Szeretnénk tisztázni a mesterséges neuronhálónk döntéseinek interpretációját, kipróbálni a tisztán a nyelvészeti szabályokkal elérhető osztályozást, illetve ezek hibridjét.

Az említett neurális hálót használó modell kipróbálható az Egészségügyi áhírfelismerő címen (URL3), illetve a kutatócsoport vendégoldalán (URL1). A működését a 2. ábra illusztrálja.



2. ábra. Az egészségügyi áhírfelismerő mesterségesintelligencia-alapú alkalmazás működés közben

Megjegyzés: A program a kijelölt helyre másolt szöveget vizsgálja. Pár másodperc alatt döntésre jut, és mutatja az eredményt, de egy sáv is jelzi az áhírnyanú százalékos mértékét.

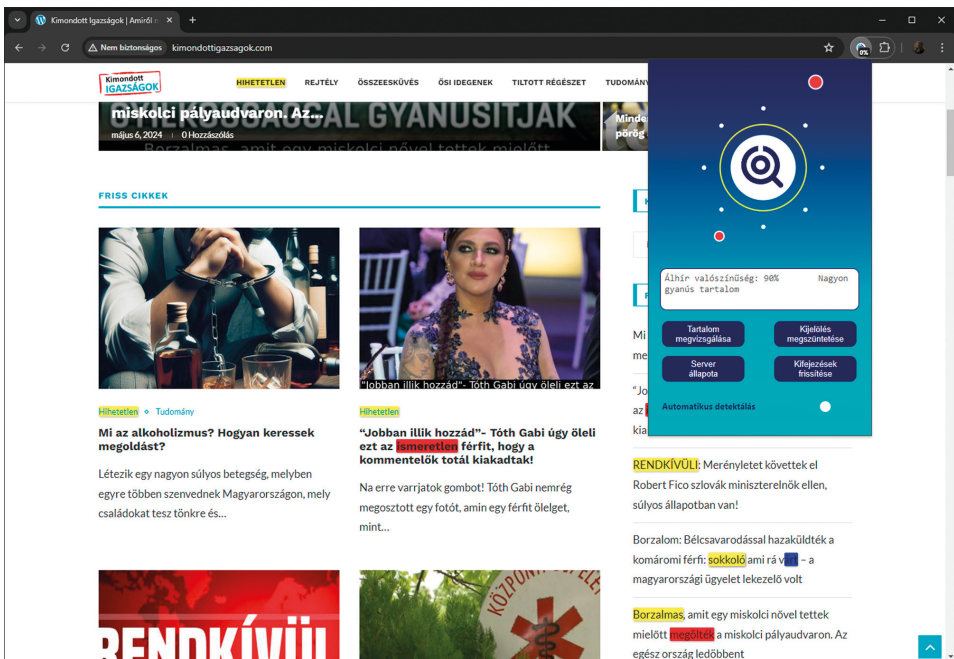
Keresőrobot

Dolgozunk egy automatikus, egészségügyi áhíreket kereső roboton is. Ez már autonóm módon, emberi beavatkozást nem igénylően lenne működőképes. Az interneten kulcsszavak segítségével keres egészségügyi szövegeket, és az előző pontban leírt algoritmussal tud áhíreket azonosítani. Ilyen eszköz fejlesztése felmerült a Nemzeti Népegészségügyi és Gyógyszerészeti Központtal (NNGYK, korábbi nevén Országos Gyógyszerészeti és Élelmszeregészségügyi Intézet, OGYÉI) tervezett együttműködésben is. A hatóság ugyan továbbra is érdekelt az együttműködésben, de az átszervezés miatt türelmünket kérték a részletek megbeszélése céljából. A keresés hatékonyságának javítása végett megkerestük az Országos Széchényi Könyvtárt (OSZK), és az SZTE-vel létrejött egy olyan együttműködési szerződés, amelynek keretében az OSZK és az SZTE a webes aratás, illetve az ehhez kapcsolódó mesterségesintelligencia-alapú

alkalmazások kifejlesztésére vonatkozó szándéknyilatkozatot fogalmaztak meg. A program rendben működik, a tesztelések a várakozásoknak megfelelő eredményeket adtak.

Keresőprogramba beépülő bővítmény

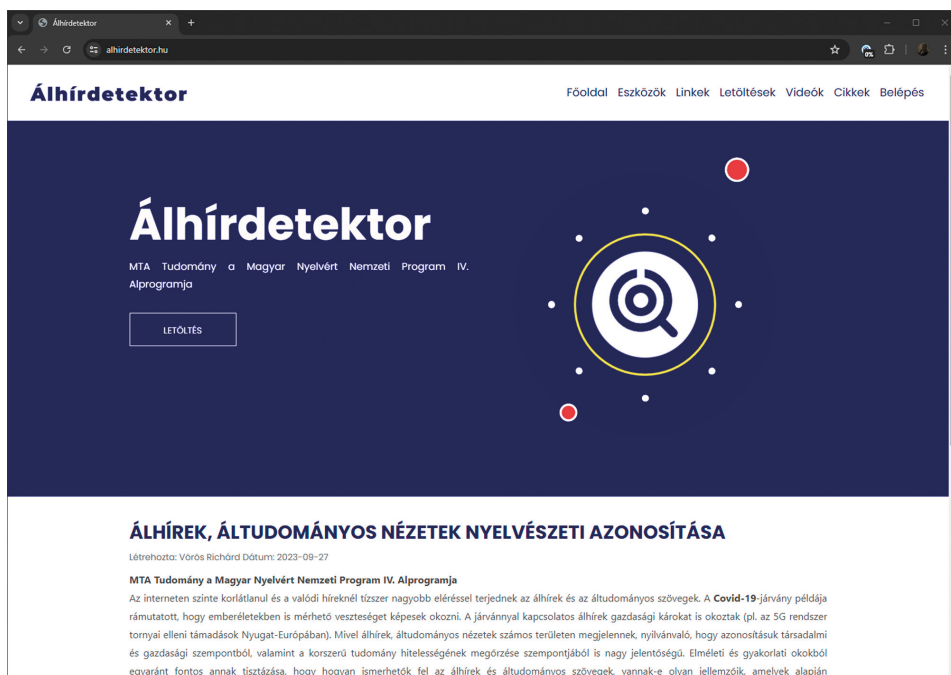
Készítettünk egy keresőprogramba beépülő álhírdetektáló alkalmazást is. Ennek továbbfejlesztésén még dolgozunk, de az első működőképes változatot bemutató videó elérhető az URL4 címen, illetve az URL1 Videók vendégoldalról. A 3. ábra illusztrálja a működését. Lényegében hasonló a tudása, mint az okostelefonos alkalmazásunké. Ennek a megoldásnak a kidolgozottsága egyelőre a legkisebb.



3. ábra. A keresőprogramba beépülő alkalmazás működés közben

Megjegyzés: Lényegében az okostelefonos alkalmazáshoz hasonló a döntése, de az oldalon található teljes szöveg alapján dönt. Az osztályozás összegzett eredményét színnel is jelzi, valamint utal a gyanús részletekre.

Folyik egy egységes vendégoldal fejlesztése is. Ezen a kutatócsoport eredményeit tesszük közzé hírekkel, letöltési lehetőséggel és bemutató anyagokkal együtt. Ennek fejlesztés alatt lévő változata már elérhető az Álhírdetektor weboldalon (URL1), a szerkeszthető hálózati környezet beüzemelése folyik még. A főoldalt mutatja a 4. ábra.



4. ábra. Az MTA–SZTE–DE Elméleti Nyelvészeti és Informatikai Kutatócsoport álhírdetektálási munkáinak gyűjtőoldala

TERVEINK

Együttműködési megállapodást kötöttünk a Pécsi Tudományegyetem (PTE) gyógyszerész kutatóival a gépi tanulás alkalmazása az online gyógyszerforgalmazói felületek webes tartalmának automatikus kategorizálását célzó kutatás közös megvalósítására. Az érdemi munka nemrég kezdődött meg. Nagy nyelvi modellen alapuló, többnyelvű mesterséges neuronhálóat fejlesztünk. Mivel a probléma megoldása az Európai Unió hatósági feladatai között szerepel, ezért keresük majd a lehetőségét, hogy a megfelelő hatóság munkáját segítsük.

Jól kiegészítjük egymást a Hungarian Digital Media Observatory (HDMO) projekttel: ők többek között tényellenőrzést végeznek, és módszertani ajánlásokat tesznek (URL5, URL6). Gyors becslést kínáló eszközeink hasznosak lehetnek számukra az előszűrésben, illetve a megkeresések egy részének kezelésében. A meglévő együttműködési szerződés alapján kölcsönös előnyök kihasználására törekszünk.

Hasonló közös munkára számítunk a Nemzeti Népegészségügyi és Gyógyszerészeti Központtal (NNGYK, korábban OGYÉI). Az első megbeszélésekre már

sor került, de az együttműködési szerződés megkötése előtt vagyunk még. Várhatóan az interneten elérhető egészségügyi tartalmú szövegek gyors előszűrése, és keresőrobot segítségével való korai figyelmeztetés lesz majd a cél. De felmerült a határozatok összeállításában segítő mesterségesintelligencia-alapú algoritmus előkészítése is.

Keressük a külföldi kutatócsoportokkal való együttműködést is, akiket hasonló kérdések foglalkoztatnak. Felvettük a kapcsolatot például a FakeSpeak projekttel (URL7). A külföldi kapcsolatok felvetik a többnyelvű rendszerek fejlesztési igényét. Aktívan szerepelünk nemzetközi konferenciákon, hogy az elért eredményeink ismertetésével bővítsük az együttműködők körét.

KÖSZÖNETNYILVÁNÍTÁS

Köszönjük az MTA támogatását a Tudomány a Magyar Nyelvért Nemzeti Program keretében, valamint a Szegedi Tudományegyetem és a Debreceni Egyetem segítségét a projekt végrehajtása helyi feltételeinek a biztosításában. Hála-sak vagyunk tanácsadóinknak, akik rendszeresen támogatják a munkánkat: Bari Ferencnek, Boldogkői Zsoltnak, Csupor Dezsőnek és Halasi Attilának, valamint az informatikai megoldások kidolgozóinak: Ficsor Tamásnak, Genci Mihálynak és Vörös Richárdnak. Sok segítséget kaptunk az SZTE Innovációs Igazgatóságától.

IRODALOM

- Berend Gábor – Csendes Tibor – Ficsor Tamás et al. (2024): Álhírek nyelvi modellek segítségével történő felismerése. *Nyelv és Tudomány*, nyest.hu, közlésre benyújtva
- Ficsor Tamás – Berend Gábor (2023): *HuBERTUSz: Alacsony paraméterszámú transzformer modellek létrehozása és kiértékelése magyar nyelvre*. XIX. Magyar Számítógépes Nyelvészeti Konferencia, MSZNY–2023. Szegedi Tudományegyetem, Szeged, 417–432.
- Nemeskey Dávid Márk (2020): *Natural Language Processing Methods for Language Modeling*. PhD-értekezés. Budapest: ELTE

URL1: Álhírdetektor, www.alhirdetektor.hu

URL2: <https://www.youtube.com/watch?v=ybg5NU-UwD4&t=36s>

URL3: Egészségügyi álhírfelismerő, <https://www.inf.u-szeged.hu/~genci/alhir-felismero.html>

URL4: https://www.alhirdetektor.hu/uploads/bongesz_bovitmeny.mp4

URL5: Magyar Digitális Média Observatórium (Lakmusz-HDMO), <https://www.hdmo.eu/section/about-us/who-we-are/?lang=hu>

URL6: Detekto, www.detekto.hu

URL7: Fakespeak, <https://www.hf.uio.no/ilos/english/research/projects/fakespeak/index.html>