FILOLÓGIA.HU • XIV. évfolyam (2023) 1–4. szám 51–64. • DOI: 10.59648/filologia.2023.1-4.3

Dodé Réka

# Magyar kulcsszavak vizsgálata és kinyerésük eredményeinek összevetése – pilotkutatás

## Research on Hungarian Keywords and Comparing the Results of their Extraction – Pilot Study

*A kulcsszó- és terminuskinyerés nem új keletű kutatási téma, már ötven éve foglalkoznak vele a kutatók, azonban még mindig rejt magában kihívásokat. A nyelvi modellek új perspektívát adnak számos nyelvtechnológiai területen, így a kulcsszó- és terminuskinyerés területén is, mivel a nyelvi modellek olyan új kulcsszavak generálására is képesek, amelyek nem, vagy csak részlegesen szerepelnek a szövegben. Amikor a szerzők kézzel adnak meg kulcsszavakat, akkor a saját háttértudásukból is merítenek, így ezek a kulcsszavak nem feltétlenül szerepelnek a szövegben. A kézzel megadott kulcsszavakkal tehát érdemes foglalkozni, és tekinthetők a gold sztenderdnek, célnak a kulcsszókinyerő alkalmazások teszteléséhez. Kutatásunkban 30 változó doménből származó tudományos szöveget és a hozzájuk tartozó szerzői kulcsszavakat vetettünk össze a ChatGPT által, többféle promptra adott megoldásokkal. Az eredmények szerint nincs szignifikáns különbség a kvantitatív eredményekben, de amennyiben kvalitatívan elemezzük a ChatGPT megoldásait, azokat relevánsnak találjuk. A dolgozat célja, hogy a ChatGPT által adott kimeneteket kiértékeljük abból a szempontból, hogy mennyire közelítik meg a szerzők által megadott kulcsszavakat.*

Kulcsszavak: terminus, kulcsszókinyerés, kulcsszómegadás, nyelvi modell, ChatGPT

*Keyword and term extraction is a well-established area of research that has attracted scholarly attention for the past five decades. However, it continues to pose persistent challenges. Language models introduce a novel dimension to various facets of natural language processing, including the realm of keyword and term extraction. They offer the capability to generate novel keywords that may be absent or only partially represented within the source text. When the authors enter keywords manually, they draw on their own background knowledge, so these keywords are not necessarily included in the text. Manually entered keywords are therefore worth dealing with and can be considered the gold standard, a benchmark, against which to test keyword extraction applications. In our study, we conducted a comparative analysis of manually assigned keywords for 30 scientifical textual documents (from different domains) against keyword solutions provided by*

*ChatGPT in response to various prompts. Our findings indicate that while there may not be a statistically significant difference in quantitative metrics, a qualitative examination of ChatGPT-generated solutions reveal their relevance and utility in augmenting keyword assignments. The aim of the thesis is to evaluate the outputs given by ChatGPT from the point of view of how close they are to the keywords given by the authors.*

Keywords: term, keyword extraction, providing keyword, language model, ChatGPT

## 1. Bevezetés

A neurális nyelvi modellek az utóbbi évtizedben új perspektívát nyitottak számos nyelvtechnológiai feladat megoldására. „A nyelvi modellek a közelgő szavak előrejelzését adják a megelőző szókontextusból" (Jurafsky – Martin 2023: 147). A korábbi nyelvi modellekkel szemben a neurális nyelvi modellek (mint amilyen a tanulmányban használt ChatGPT is) „sokkal hosszabb előzményeket tudnak kezelni, jobban általánosíthatók a hasonló szavak kontextusában, és pontosabbak a szó előrejelzésében" (Jurafsky – Martin 2023: 147).

A transzformátor olyan nyelvi modell, amely új mechanizmusokat kínál, és ezáltal hatékonyabban kezeli az egymástól távolabb eső szavak kapcsolódását (Vaswani et al. 2017). A terminológiai munkafolyamat megkönnyítése is célja lehet a nyelvi modellek alkalmazásának. Az elmúlt évben többen foglalkoztak azzal, hogy a nagy nyelvi modelleket finomhangolják terminológiai munkára, például terminuskinyerésre (Gu et al. 2021; Vakili et al. 2022; Zheng et al. 2022), illetve hogy ChatGPT-lekérdezéseket (2023) használjanak a terminológiai munka megkönnyítésére (például Pantcheva 2023; Smullen 2023).

A finomhangolás az a folyamat, amelynek során az előtanított modellt továbbtanítják valamilyen feladatra (Jurafsky – Martin 2023). Ilyen feladatok lehetnek például a koreferencia feloldása vagy a névelemek felismerése a szövegben. Előtanítás során egy modell nagyon nagy mennyiségű szöveg feldolgozásával a szavak vagy mondatok jelentésének valamiféle reprezentációját tanulja meg (Jurafsky – Martin 2023).

A kulcsszavak azok a nyelvi elemek, amelyek reprezentálják a szöveget, betekintést nyújtanak a szöveg tartalmába, továbbá segítik a könnyebb eligazodást a szakszövegek között (Berend – Farkas 2010).

„A terminus olyan elnevezés, amely nyelvi eszközökkel egy általános fogalmat reprezentál (…) a terminológia egy doménhez vagy tárgykörhöz tartozó megnevezések és fogalmak összessége" (ISO 1087: 2019, ford.: a szerző).

A kulcsszókinyerés és a terminuskinyerés összeérnek több ponton, ahogy azt Nomoto (2023) is írja: „a terminológiatudomány és az információkinyerés az idők során egymás mellett haladtak a fejlődésben, és néha keresztezték egymás útját. A terminológusok olyan kifejezéseket keresnek, amelyek egy doménra jellemzőek, és amelyek hasznosak az adott területtel kapcsolatos ismeretek rendszerezéséhez, míg az információkinyerésben a kifejezések azonosítására összpontosítanak (ezeket indexelési terminusnak nevezik), amelyek képesek megkülönböztetni a dokumentumokat" (Nomoto 2023).

A kulcsszókinyerést lehet osztályozási problémaként kezelni (Firoozeh et al. 2019), ebben az esetben a tanítóanyag kulcsszavai egy nagyszámú vagy nyitott végű osztály-, címkekészlet elemei. Ez a felügyelt gépi tanulás egy típusa. A felügyelt gépi tanulás esetében tanító adatbázist használunk, amely a tapasztalatokat, megfigyeléseket tartalmazza (ez esetben a lehetséges címkéket). Ez lesz az elvárt célérték, amely alapján egy olyan modell készül, amely a nem látott példákon is helyesen működik (Farkas é. n.). Erre egy példa a magyar nyelvű kulcsszó-, címkekinyerő, dokumentumosztályozó is, amelyet Yang és munkatársai finomhangoltak magyar nyelvre a HVG szövegeivel (Yang et al. 2020).

Nomoto (2023) összegezve az elmúlt ötven év kulcsszókinyerési alkalmazásait és megoldásait, arra a következtetésre jut, hogy a hatékonyság fontossága miatt a felügyelet nélküli módszerek irányába kellene továbbmenni (gyorsabban futnak, kevesebb erőforrást igényelnek), még ha ez a pontosság rovására is megy, illetve áttérni a generatív rendszerre, mivel az lehetőséget ad olyan kulcsszavak generálására is, amelyek nem, vagy csak részlegesen szerepelnek a szövegben.

A ChatGPT (Chat Generative Pre-trained Transformer) egy nyelvi modell, amelyet az OpenAI fejlesztett, és a transzformátor-architektúrán alapul. A ChatGPT esetében a modellt kérdések megválaszolására és beszélgetéskezelésre finomhangolták egy kisebb, feladatspecifikus adatkészlettel (Mottesi 2023). A ChatGPT képes olyan szövegeket is létrehozni, amelyek nem, vagy csak részben szerepelnek a tanító anyagban.

A ChatGPT hátránya, hogy maximum 4096 tokenből (szövegszóból) álló szöveget lehet beadni promptként (a kutatás ideje alatt). A prompt (felszólítás) az az utasítás/kérés vagy kérdés, amelynek hatására kimenetet kapunk az erre finomhangolt modelltől (Jalalov – Gaszcz 2023). A prompt mennyiségének korlátait kívánja orvosolni a LlamaIndex (egy adatfeldolgozó keretrendszer, amely képes a GPT-modellekkel kommunikálni), amely a nagyobb méretű szöveget feldarabolja, és úgy indexálja (LlamaIndex 2023). Jelen kutatásba a LlamaIndexet nem vettük bele, azt a későbbiekben tervezzük tesztelni hosszabb szövegeken, hogy kiaknázhassuk ezt a képességét.

Kutatásunkban 30 különböző doménből származó tudományos szöveget és a hozzájuk tartozó manuálisan megadott kulcsszavakat hasonlítottuk össze kvantitatív és kvalitatív módon a ChatGPT által többféle promptra adott kimenetekkel. Célunk kiértékelni a ChatGPT által adott kimeneteket abból a szempontból, hogy mennyire közelítik meg a szerzői kulcsszavakat. Ezek alapján hipotéziseink a következők: 1. a ChatGPT a legjobb eredményt akkor éri el, amikor a kulcsszót és terminust egyszerre kérő promptot kapja a modell; 2. a legrosszabb eredményt akkor adja, amikor a csak kulcsszót kérő promptot kapja a modell.

### *1.1. Motiváció és előzmények*

A terminuskinyerés a terminológiamenedzsment (a teljes terminológiai munkafolyamatra vonatkozó fogalom; Tamás 2014) szerves része, amely – akár csak manuálisan, akár (fél)automatikusan, majd manuálisan végezzük – fontos és időigényes feladat. Amikor felügyelt gépi tanulás alapú terminuskinyerő alkalmazásokat fejlesztünk, elengedhetetlen egy terminuslista megléte, amelyre az automatizált gépi tanulás

módszereit alkalmazzuk. Ha elérhetők külső források (terminológiai adatbázis vagy szólista) a doménspecifikus terminológia számára, akkor azok használhatók (vö. Mihalcea – Csomai 2007), amennyiben nem áll ilyen rendelkezésre, akkor a tanító adatokat manuálisan kell összeállítani.

A terminuskinyerés összetettebb feladat, mint a kulcsszókinyerés, mivel a terminusok generálását is magában foglalja, egyrészt mivel a technológia fejlődésével a szakszókincs is folyamatosan bővül, változik, másrészt a terminusok nem minden esetben fordulnak elő az adott szövegben, ahogy a kézzel megadott kulcsszavak sem feltétlenül fordulnak elő a korpuszban. Hulth (2003) is rámutat, hogy a szerzők saját háttértudásukból is adnak meg kulcsszavakat, és azok nem feltétlenül szerepelnek a szövegben, illetve nem feltétlenül gyakoriak, mint a statisztikai (gyakoriságon alapuló) módszerekkel kinyert kulcsszavak. Erre a következtetésre jut Dodé (2023) is, aki a kézzel megadott kulcsszavak szövegbeni előfordulását vizsgálta. Rámutatott, hogy az általa vizsgált korpuszban a kulcsszavak 37%-a kevesebb mint kétszer fordul elő a szövegekben. Ezt azonban nem csupán elírások vagy morfoszintaktikai eltérések okozzák, hanem a szerzők tudatosan általánosabb vagy speciálisabb kulcsszót használnak, illetve figyelembe veszik, hogy a szöveget szintén a szakma ismerői fogják olvasni (Dodé 2023). Ilyen értelemben maga a terminuskinyerés terminus nem megfelelő, mivel a kinyerés szó azt sugallja, hogy a meglévő szövegből nyeri ki a lehetséges terminusokat az alkalmazás.

A kulcsszavak, amelyeket a szerző ad hozzá a szöveghez, a leghatékonyabban reprezentálják a szöveg tartalmát, mivel a kulcsszóadási stratégia nem szövegbeli gyakoriságon alapul, illetve mert nem ragaszkodik a szöveg szókincséhez, továbbá mivel a szöveg szerzője a szakmai tudás birtokában adja meg ezeket a kifejezéseket. Ezek alapján a kifejezések alapján tudjuk a legbehatóbban feltérképezni az adott szöveg fogalmi rendszerét. „A fogalmi rendszer a fogalmak összefüggő, bizonyos szempontok szerint rendezett hálózata" (Arntz et al. 2009, idézi Tamás 2014: 34), „más néven fogalmi háló, és célja, hogy bemutassa, hogyan viszonyulnak egymáshoz a fogalmak" (Tamás 2014: 34). A fogalmi rendszer által könnyebb megérteni, definiálni egy fogalmat. A fogalmi rendszert használják a fogalmi harmonizációhoz, amely fontos alfeladata a terminológiamenedzsmentnek. A fogalmi harmonizáció „olyan tevékenység, amelynek célja a két vagy több, egymással szorosan összefüggő vagy egymást átfedő fogalom közötti különbségek megszüntetése vagy csökkentése" (ISO 860: 1). A kézzel megadott kulcsszavak tehát kiemelten fontosak, érdemes foglalkozni velük és tanítóanyagként használni őket.

A jelen kutatásban tudományos szövegeket vizsgálunk, amelyek esetében a szerzői kulcsszavakra elvétve található instrukció. Leggyakrabban a megadható kulcsszavak számát találjuk szerzői útmutatóban, de van, ahol orientálja is a szerzőket a kiadvány. A Springer Kiadó (Springer é. n.) irányelvei kiemelik, hogy a kulcsszavak megadása olyan eszköz, amely segíti az indexelést és a keresőmotorokat a releváns cikkek megtalálásában, a kulcsszavak pedig a tartalmat reprezentálják, és meghatározzák a szakterületet. Az útmutató példákat ad a jó és a rossz kulcsszavakra (általában hangsúlyozva, hogy egy kulcsszó akkor jó, ha elég konkrét, például *climate change, erosion* vs. *quaternary climate change, soil erosion*). E motivációt követve a kutatásban a szövegeket és a kézzel megadott kulcsszavaikat tekintjük a gold sztenderdnek (referenciaértékek készlete). A gold sztenderd címkékkel hasonlítjuk össze a modell előrejelzéseit a kiértékelés során (Farkas é. n.).

A dolgozat célja, hogy kiértékeljük a ChatGPT által adott kimeneteket abból a szempontból, hogy mennyire közelítik meg a szerzők által megadott kulcsszavakat.

## 2. Eszközök és módszertan

A vizsgálat során a ChatGPT 3.5-ös modelljét[1] használtuk, amelynek 4 kérdést tettünk fel.

### *2.1. A vizsgált szövegek*

A 30 szövegből álló korpuszt a REAL repozitórium cikkeiből állítottuk össze, amelyek egy része mindenki számára böngészhető és letölthető. A REAL az MTA Könyvtár és Információs Központ repozitóriuma, amely a „hazai szakcikkek és kutatási jelentések tárolására" készült (REAL).

A szövegeket háromféle szűrővel dolgoztuk fel: 1. a főszöveg nyelvének meghatározásával (Python nyelvdetektáló segítségével) biztosítottuk, hogy a szövegek magyar nyelvűek legyenek; 2. a szövegekben a manuálisan megadott kulcsszavak jelenlétét ellenőriztük; 3. azokat a szövegeket választottuk ki, amelyek mérete nem haladta meg a 25 kilobyte-ot, figyelembe véve a ChatGPT tokenkorlátjait. Eme lépések eredményeként 203 szöveget kaptunk, amelyekből 30 véletlenszerűen kiválasztott szöveget használtunk a kutatásunkhoz.

Ezek a szövegek összesen 142 kézzel meghatározott kulcsszót tartalmaznak. Átlagosan 4,7 kulcsszó szerepel szövegenként, maximum 10, minimum 2 kulcsszó fordul elő egy szövegben. A leggyakoribb esetben, összesen 9 alkalommal, 4 darab kulcsszó található meg egy adott szövegben. Ezek a kulcsszó-szöveg párosok képezik a vizsgálatunk gold sztenderd korpuszát.

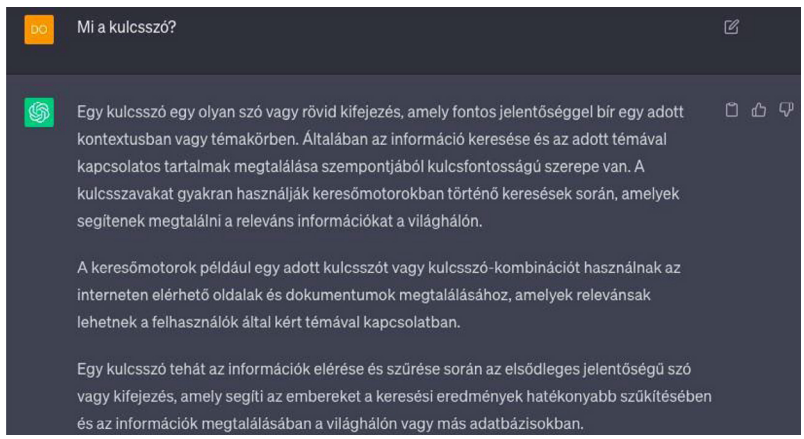
### *2.2. ChatGPT*

A chat.openai.com webalkalmazás tesztelése előtt az esetlegesen fennmaradt angol nyelvű kulcsszavakat eltávolítottuk a szövegekből, valamint javítottuk az elválasztásokat, és eltávolítottuk a sortöréseket annak érdekében, hogy értelmezhetőbb legyen a szöveg, és hogy csökkentsük a tokenszámot. Ezt követően a modellnek négy promptot (instrukciót, felszólítást) adtunk, amelyek a következők voltak:

1. Sorold fel a következő szöveghez tartozó X olyan kifejezést, amely kulcsszó és terminológiai fogalom is: [szöveg]
2. Sorold fel a következő szöveghez tartozó X kulcsszót: [szöveg]
3. Sorold fel a következő szöveghez tartozó X terminust (terminológiai fogalom): [szöveg]
4. Sorold fel a következő szöveghez tartozó X terminust. Lehetnek tulajdonnevek is: [szöveg]

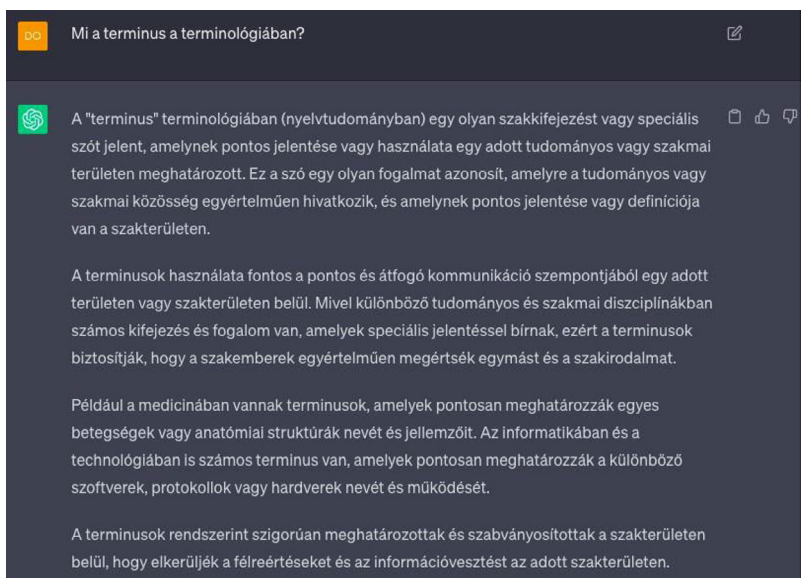
1 platform.openai.com/docs/guides/gpt

A lekérdezés előtt kíváncsiak voltunk, hogyan definiálja a ChatGPT a kulcsszót (1. ábra), a terminust (2. ábra), és a kettő között szerinte mi a különbség (3. ábra). A válaszokat a következő ábrák mutatják. Ezt követően adtuk be a modellnek a szövegeket és a fenti promptokat felhasználva a modell memóriáját.

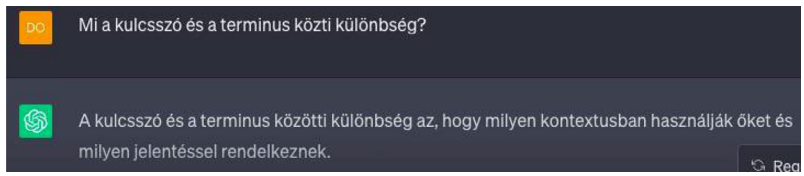



*1. ábra: A „Mi a kulcsszó?" kérdésre adott válasz (ChatGPT)*
*Forrás: a szerző szerkesztése*

A terminus több doménben is használt szakkifejezés, ezért a terminológia domént konkretizáltuk a kérdésben.




*2. ábra: A „Mi a terminus?" kérdésre adott válasz (ChatGPT)*
*Forrás: a szerző szerkesztése*




*3. ábra: A „Mi a kulcsszó és a terminus közti különbség?" kérdésre adott válasz (ChatGPT)*
*Forrás: a szerző szerkesztése*

A ChatGPT angol nyelvű anyagokon volt előtanítva, emiatt előfordul, hogy angol választ generál. Ahelyett azonban, hogy a lekérdezéseinknél kifejezetten rögzítettük volna, hogy magyar nyelven várjuk a választ, magyar nyelvet használtunk a kommunikáció során, ezzel aktiválva a megfelelő, magyar nyelvi válaszokat. Habár előfordult néhányszor, hogy a rendszer angolul válaszolt, az ismételt kérdésfeltevés a kívánt válaszokat eredményezte.

A *sorold* szót alkalmaztuk annak érdekében, hogy a ChatGPT kerülje a definíciókat. Ennek ellenére gyakran előfordult, hogy a terminusra vonatkozó promptokra definíciót és angol ekvivalenst is hozzáfűzött a válaszaihoz. A definíciókat ignoráltuk.

A kért kulcsszavak és terminusok száma dinamikusan változott attól függően, hány kézzel meghatározott kulcsszó volt jelen az egyes szövegekben. Amikor például három kulcsszó szerepelt, akkor három kulcsszót és terminust kértünk, de amikor hét kulcsszó volt jelen, akkor hetet. Néhányszor előfordult, hogy a válaszban hibákat ejtett a rendszer ezen a téren, különösen akkor, amikor kulcsszavakat kértünk, és többet adott, mint amennyit vártunk. Volt olyan eset, amikor további magyarázatot fűzött a válaszhoz. Például: „Kérlek, vedd figyelembe, hogy ezek a terminusok a szövegben található információk alapján kerültek meghatározásra, és kontextuson kívül eltérő jelentéssel is rendelkezhetnek" (ChatGPT). Ezt szintén ignoráltuk.

A kézzel megadott kulcsszavak a korábbi kutatás (Dodé 2023) szerint bizonyos szempontból tekinthetők terminusoknak, így a lekérdezésnél nemcsak a kulcsszavakra kérdeztünk rá, hanem a terminusokra is. A 4. kérdésben pedig hozzátettük, hogy tulajdonneveket is adhat, mivel a Dodé által vizsgált anyagban a kézzel megadott kulcsszavak 27%-a tulajdonnév volt (Dodé 2023).

## 3. Eredmények

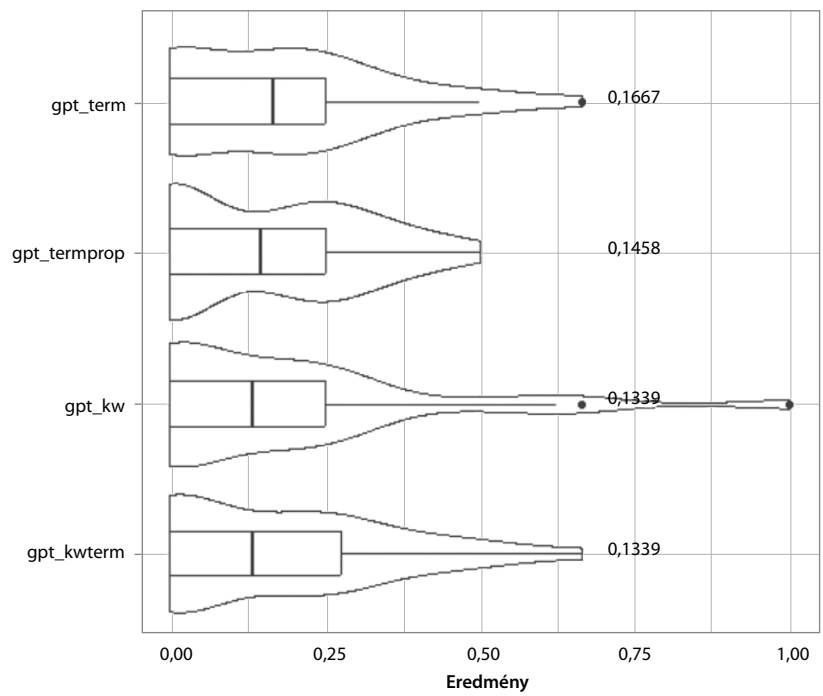
A kvantitatív eredményeket úgy értékeltük, hogy a megoldásokat két kategóriába osztottuk:

1. egyező = verbatim egyezés,
2. majdnem egyező = eltérő írásmód, más szófaj (például *laparoszkópia* vs. *laparoszkópos),* eltérő toldalékkal adott megoldás (például *fosszilis energiák* vs. *fosszilis energia*), vagy angol ekvivalenst is tartalmaz (például *innováció* vs. *innováció [innovation]*).

Minden lekérdezés eredményét összegeztük, figyelembe véve az egyező és a majdnem egyező találatokat, majd ezt az összeget elosztottuk a szöveghez tartozó, kézzel megadott kulcsszavak számával. A második kategória eredményeit 0,5-tel szoroztuk meg annak érdekében, hogy tükrözze a teljes egyezés és a majdnem egyezés közötti különbséget. Az így kapott mutatók 0 és 1 közötti értékeket vehetnek fel, ahol az érték 1, ha a lekérdezés az összes kulcsszót helyesen azonosította, és 0, ha egyetlen kulcsszóval sem talált egyezést.

### *3.1. Kvantitatív eredmények*

A 4. ábrán bemutatott eredmények szóródása jól látható a különböző lekérdezési típusok között. Az interkvartilis tartomány minden esetben közel 0,25 körül helyezkedik el. Érdekesség, hogy a ChatGPT a csak kulcsszókérő lekérdezéseknél mutat kiemelkedő szélsőértékeket. Egyik esetben sem tapasztalható normál eloszlás. Az ábrán látható számok a medián értékeit jelölik, mivel ezek kevésbé érzékenyek a szélsőértékekre. A legmagasabb medián érték a csak terminuskéréseket tartalmazó lekérdezés esetében figyelhető meg (0,1667).




*4. ábra: Az eredmények szövegenkénti szóródása*
*Forrás: a szerző szerkesztése*

Az eredményeket az 1. táblázatban is láthatjuk.

*1. táblázat: Az eredmények mediánja, átlaga és a promptok legjobb eredményei (saját szerkesztés)*

| | kwterm | kw | term | termprop |
|---|---|---|---|---|
| Medián | 0,1339 | 0,1393 | **0,1667** | 0,1458 |
| Átlag | 0,1721 | **0,1848** | 0,1798 | 0,1540 |
| Legjobb eredmény (max) | 0,6667 | **1,0000** | 0,6667 | 0,5000 |

*Forrás: a szerző szerkesztése*

A táblázatban háromféle adatsort találunk: a prompttípusok (kulcsszó + terminus = kwterm; kulcsszó = kw; terminus = term; terminus + tulajdonnév = termprop) mediánját, átlagát és legjobb eredményét. Ha alaposabban megvizsgáljuk ezeket a számokat, a következő különbségek rajzolódnak ki: az átlagot tekintve a ChatGPT akkor érte el a legjobb eredményt, amikor csak kulcsszót kértünk tőle a promptban, míg a medián alapján a terminuskérő lekérdezés teljesített a legjobban. Az is látszik, hogy a kulcsszókérő lekérdezés az egyik szöveg esetében az összes kulcsszót megtalálta.

Az összesített eredmények alapján azt láthatjuk, hogy a kézzel megadott kulcsszavak és a ChatGPT megoldásai ritkán egyeznek meg – az átlagosan elért legjobb eredmény mindössze 0,1848 (medián: 0,1667).

A Kruskal–Wallis-próba alapján nincs szignifikáns eltérés a különböző lekérdezési típusok eredményei között, így statisztikailag egyik hipotézisünk sem igazolódott.

## *3.2. További megfigyelések*

Annak ellenére, hogy a kézzel megadott kulcsszavak között gyakran szerepelnek tulajdonnevek (Dodé 2023), megfigyelhető, hogy a ChatGPT nem mindig volt képes megfelelően értelmezni a kérdést. Más esetekben viszont épp ellenkezőleg, túlteljesítette a feladatot, ahogy azt a 2. táblázat szemlélteti.

*2. táblázat: Példa a terminus és a tulajdonnév promptra adott eredményből*

| **Kézzel megadott kulcsszavak** | | | | | |
|---|---|---|---|---|---|
| terminológia mesterszak | terminológia | terminográfiai módszerek és eszközök | károli gáspár református egyetem | tudás-menedzsment | információ-kezelés |
| **Terminus és tulajdonnév** | | | | | |
| B. Papp Eszter | Fóris Ágota | Bölcskei Andrea | Terminográfiai módszerek és eszközök a terminológus-képzésben | Károli Gáspár Református Egyetem, BTK, TERMIK, 1088 Budapest, Reviczky u. 4. | Terminológia mesterszak |

*Forrás: a szerző szerkesztése*

Több szövegnél hasonló eredmények jöttek ki (3. táblázat). Ebben a példában ugyanazok a kulcsszavak (*Aortabillentyű-sebészet*, *Ministernotomia*) szerepelnek, azonban ezek nem felelnek meg a kézzel megadott kulcsszavaknak.

*3. táblázat: Példa a hasonló eredményekre*

| | | |
|---|---|---|
| **kézzel** | minimális behatolással végzett műtét | aortabillentyű-csere |
| **term + kw** | Ministernotomia | Aortabillentyű-sebészet |
| **kw** | varrat nélküli billentyűk | Aortabillentyű-sebészet |
| **term** | Ministernotomia | Aortabillentyű-sebészet |
| **term + prop** | Ministernotomia | Aortabillentyű-sebészet |

*Forrás: a szerző szerkesztése*

Az *aortabillentyű-sebészet* az *aortabillentyű-csere* fölött lévő, generikusabb fogalom. A *ministernotomia* pedig egy műtéti megoldás (Szabolcs 2011). Ebben az esetben a szerző azt a döntést hozta, hogy a magyar nyelvű, kevésbé szakmai, magyarázó kulcsszavakat alkalmazza.

Előfordul olyan példa is, ahol több kulcsszó közül néhány azonos a kézzel megadott kulcsszavakkal, ahogy azt a 4. táblázat is illusztrálja.

*4. táblázat: Példa az egyező eredményekre*

| | | | | |
|---|---|---|---|---|
| **kézzel** | desmodesmus communis | nehézfém-szennyezés | réz | biológiai fémmentesítés |
| **term + kw** | Hidrológiai Közlöny | Desmodesmus communis | Réz-akkumuláció | Rézeltávolítás |
| **kw** | Hidrológiai Közlöny | Desmodesmus communis | Réz-akkumuláció | Érzékenység |
| **term** | Hidrobiológia | Desmodesmus communis | Réz-akkumuláció | Rézeltávolítás |
| **term + prop** | Hidrológiai Közlöny | Desmodesmus communis | Réz-akkumuláció | Rézeltávolítás |

*Forrás: a szerző szerkesztése*

Itt a szerző is a latin *desmodesmus communis* terminust adta meg kulcsszóként, azonban általános fogalmat is használt (*réz*), amellyel nem volt egyezés.

Az eredmények, ha csak a számszerű értékeket vesszük figyelembe, nem tekinthetők kielégítőnek. Azonban sok esetben megfigyelhető, hogy a válaszok alapvetően értelmezhetők és relevánsak voltak. Több esetben az eltérések oka a szerző által használt fogalmak más szintjére vezethető vissza; generikus vagy specifikus fogalmak használata. Például: *csont* vs. *állkapocscsont* (specifikusabb); *Magyarország flórája* (specifikusabb) vs. *flóra*; *koleszterin* vs. *koleszterinszint* (szinonima); *xenorhabdus budapestensis* (specifikusabb) vs. *budapestensis*.

## 4. Következtetések

Kutatásunk célja az volt, hogy 30 szöveget és a hozzájuk kézzel megadott kulcsszavakat, amelyek gold sztenderdnek tekinthetők, összevessük a ChatGPT különböző promptjai által kapott eredményekkel. Összességében láthatjuk, hogy a kulcsszóadási stratégia eltér a szerző és a ChatGPT esetében. Ha a kvantitatív eredményekre koncentrálunk, láthatjuk, hogy a nyelvi modell nem teljesített a hozzá társított várakozásoknak megfelelően, mivel a legjobb medián is csak 0,1667 volt (a maximális érték 1,0 lehetett). A Kruskal–Wallis-próba eredménye alapján pedig nincs szignifikáns eltérés a különböző lekérdezési típusok eredményei között. Az elért eredmények háttérében azonban több tényező is szerepelhet:

1. A ChatGPT korlátozott tokenszámára (maximalizált beadható tokenszám) való tekintettel, a vizsgált anyagok a tanulmányokhoz képest rövid szövegek, így a hozzájuk tartozó kulcsszavak megadása is nagyobb kihívás. Erre megoldást nyújthatnak az olyan keretrendszerek, amelyek segítségével nagyobb méretű szöveget adhatunk a ChatGPT-nek (például LlamaIndex), akár darabolva is, de anélkül, hogy információt veszítenénk.
2. Ezenkívül a modell korlátozott mozgásterét is érdemes megemlíteni; a kulcsszókérési limit növelése, például az első 15 kulcsszóra történő korlátozás az eddigi 2–10 helyett, javíthatja az egyezéseket.
3. Habár a szövegek OCR-feldolgozása során javításokat eszközöltünk, még mindig előfordulhatnak hibák, amelyek negatívan befolyásolhatják az eredményeket.
4. Emellett a lekérdezések átfogalmazása és új instrukciók megfogalmazása is segíthet a válaszok minőségének javításában, de egzakt, mindig megegyező választ nem feltétlenül fogunk kapni.

Egy következő tanulmányban tervezzük megvizsgálni a ChatGPT kreativitását: a kapott eredmények közül van-e olyan, amely nincs a szövegben – összevetve a kézzel megadott, szövegben nem szereplő kulcsszavakkal –, tehát generált-e új kulcsszót a nyelvi modell, illetve ha igen, milyen stratégia mentén (amennyiben többféle stratégiát használt, azok kategorizálása). Ez egyéb kapaszkodókat adhat a nyelvi modell működésének megértéséhez. Noha a kézzel megadott kulcsszavak reprezentálják a legjobban az adott szöveget, és ettől a kvantitatív eredményeket tekintve eltér a ChatGPT, a megoldásai értelmezhetők és relevánsak. Sok esetben az eltérések oka a szerző által használt fogalmak más szintjére vezethető vissza (szinominák, többelemű terminusok használata). Érdemes tehát megfontolnunk egy új, eme típusú eredményeket célzó validálási stratégia kidolgozását.

## Szakirodalom

Arntz, Reiner – Picht, Heribert – Mayer, Felix 2009: *Einführung in die Terminologiearbeit*. 6. Aufl. Hildesheim – Zürich – New York: Georg Olms Verlag.

Berend Gábor – Farkas Richárd 2010: Kulcsszókinyerés magyar nyelvű tudományos publikációkból. In: Tanács Attila – Vincze Veronika (szerk.): *VII. Magyar Számítógépes Nyelvészeti Konferencia. Szeged, Magyarország, 2010. december 2–3*. Szeged: Szegedi Tudományegyetem Informatikai Tanszékcsoport. 47–55.

Dodé Réka 2023: *Kulcsszavak és terminusok vizsgálata a REAL repozitóriumának anyagán – pilot kutatás.* Előadás. Tudásmegosztás, információkezelés, alkalmazhatóság. XXIX. Magyar Alkalmazott Nyelvészeti Kongresszus, 2023. március 17–18. Budapest: Szaknyelvi Intézet, Semmelweis Egyetem.

Farkas Richárd é. n.: *Gépi tanulás a gyakorlatban*. *Gépi tanulás alapfogalmai.* Online: www.inf.u-szeged.hu/~rfarkas/ML20/alapfogalmak.html

Firoozeh, Nazanin – Nazarenko, Adeline – Alizon, Fabrice – Daille, Béatrice 2020: Keyword Extraction: Issues and Methods. *Natural Language Engineering* 26/3: 1–33. Online: https://doi.org/10.1017/S1351324919000457

Gu, Yu – Tinn, Robert – Cheng, Hao – Lucas, Michael – Usuyama, Naoto – Liu, Xiaodong – Naumann, Tristan – Gao, Jianfeng – Poon, Hoifung 2021: Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing. *ACM Transactions on Computing for Healthcare* 3/1: 1–23. Online: https://doi.org/10.1145/3458754

Hulth, Anette 2003: Improved Automatic Keyword Extraction Given More Linguistic Knowledge.. In: Collins, Michael – Steedman, Mark (eds.): *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing.* Barcelona, Spain: Association for Computational Linguistics. 216–223. Online: https://doi.org/10.3115/1119355.1119383

ISO 860 = ISO Central Secretary 2007: *Terminology Work – Harmonization of Concepts and Terms.* Geneva, CH: International Organization for Standardization. Online: www.iso.org/standard/40130.html

ISO 1087 = ISO Central Secretary 2019: *Terminology Work and terminology science – Vocabulary.* Geneva, CH: International Organization for Standardization. Online: www.iso.org/standard/62330.html

Jalalov, Damir – Gaszcz, Karolina 2023: Legjobb Prompt Engineering Ultimate Guide 2023: Kezdőtől haladóig. *Metaverse Post,* 2023. május 14. Online: https://mpost.io/hu/prompt-engineering-ultimate-guide/

Jurafsky, Dan – H. Martin, James 2023: *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Third Edition draft. Online: https://web.stanford.edu/~jurafsky/slp3/ed3book_jan72023.pdf

Mihalcea, Rada – Csomai, András 2007: Wikify!: Linking Documents to Encyclopedic Knowledge. In: J. Silva, Mário – A. F. Laender, Alberto – Baeza-Yates, Ricardo – L. McGuinness, Deborah – Olstad, Bjorn – Haug Olsen, Øystein – O. Falcão, André

(eds.): *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management.* New York, NY, United States: Association for Computing Machinery. 233–242. Online: https://doi.org/10.1145/1321440.1321475

Mottesi, Celeste 2023: What is ChatGPT? An Introduction to OpenAI's Conversational AI Model. *InvGate,* 2023. február 2. Online: https://blog.invgate.com/what-is-chatgpt

Nomoto, Tadashi 2023: Keyword Extraction: A Modern Perspective. *SN Computer Science,* 92/4. Online: https://doi.org/10.1007/s42979-022-01481-7

Pantcheva, Marina 2023: Terminology Management Made Easier with Large Language Models. *RWS Blog*, 2023. május 18. Online: www.rws.com/blog/terminology-management-made-easier-with-large-language-model s/

Smullen, Daniel 2023: How To Use ChatGPT For Keyword Research. *Search Engine Journal*, 2023. április 19. Online: www.searchenginejournal.com/ChatGPT-for-keyword-research/483848/

Springer = sz. n. é. n.: Title, Abstract and Keywords. The Importance of Titles. *Springer*. Online: www.springer.com/kr/authors-editors/authorandreviewertutorials/writing-a-journal-ma nuscript/title-abstract-and-keywords/10285522

Szabolcs Zoltán 2011: *Sternotomia*. 2011. augusztus 28. Online: www.szabolcszoltan.hu/patients/sternotomy.php

Tamás Dóra Mária 2014: *Gazdasági szakszövegek fordításának terminológiai kérdései*. Budapest: ELTE Eötvös Kiadó.

Vakili, Thomas – Lamproudis, Anastasios – Henriksson, Aron – Dalianis, Hercules 2022: Downstream Task Performance of BERT Models Pre-Trained Using Automatically DeIdentified Clinical Data. In: Calzolari, Nicoletta – Béchet, Frédéric – Blache, Philippe – Choukri, Khalid – Cieri, Christopher – Declerck, Thierry – Goggi, Sara – Isahara, Hitoshi – Maegaard, Bente – Mariani, Joseph – Mazo, Hélène – Odijk, Jan – Piperidis, Stelios (eds.): *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association. 4245–4252.

Vaswani, Ashish – Shazeer, Noam – Parmar, Niki – Uszkoreit, Jakob – Jones, Llion – N. Gomez, Aidan – Kaiser, Lukasz – Polosukhin, Illia 2017: Attention Is All You Need. *Cornell University arXiv,* 2017. június 12. Online: https://arxiv.org/abs/1706.03762

Yang Zijian Győző – Novák Attila – Laki László János 2020: Automatic Tag Recommendation for News Articles. In: Kovásznai Gergely – Fazekas István – Tómács Tibor (eds.): *Proceedings of the 11th International Conference on Applied Informatics (ICAI 2020)*, Eger, Hungary, January 29–31, 2020. Volume 2650 of CEUR Workshop Proceedings. CEUR-WS.org. 442–451. Online: https://ceur-ws.org/Vol-2650/paper45.pdf

Zheng, Zhe – Lu, Xin-Zheng – Chen, Ke-Yin – Zhou, Yu-Cheng – Lin, Jia-Rui 2022: Pretrained Domainspecific Language Model for General Information Retrieval Tasks in the AEC Domain. *Cornell University arXiv,* 2022. március 9. Online: https://arxiv.org/abs/2203.04729

## Források

ChatGPT = *ChatGPT*. https://chat.openai.com/
REAL = *REAL Repozitórium*. http://real.mtak.hu/
LlamaIndex = *LlamaIndex*. www.llamaindex.ai/

**Dodé Réka** alkalmazott nyelvész, a HUN-REN Nyelvtudományi Kutatóközpont Nyelvtechnológiai kutatócsoportjának és a Terminológiai kutatócsoportjának tudományos segédmunkatársa. Az ELTE BTK Nyelvtudományi Doktori Iskola Alkalmazott nyelvészet oktatási programján abszolvált. Kutatási terület: A tárgyszavak, kulcsszavak és terminusok vizsgálata terminológiai és nyelvtechnológiai szempontból. Korpuszépítés és a terminusjelöltek kinyerése gépi tanulással nyelvi modellek segítségével. E-mail: dode.reka@nytud.hun-ren.hu