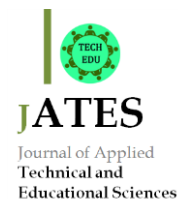




<http://jates.org>

# Journal of Applied Technical and Educational Sciences jATES

ISSN 2560-5429



## Business applications of natural language processing

Zsolt Krutilla<sup>1,2</sup>, Attila Kovari<sup>3,4,5,6</sup>

<sup>1</sup> Department of Software Development and Applications, Institute of Computer Engineering, University of Dunaujváros, Táncsics Mihály street 1/A, Dunaujváros 2401, Hungary, [krutillazs@uniduna.hu](mailto:krutillazs@uniduna.hu)

<sup>2</sup> Doctoral School of Applied Informatics and Applied Mathematics, Óbuda University, Bécsi street 96/B, Budapest 1034, Hungary

<sup>3</sup> Institute of Digital Technology, Faculty of Informatics, Eszterházy Károly Catholic University, Eszterházy tér 1, 3300 Eger, Hungary [kovari.attila@uni-eszterhazy.hu](mailto:kovari.attila@uni-eszterhazy.hu)

<sup>4</sup> Department of Natural Sciences, Institute of Engineering, University of Dunaujváros, Táncsics M. u. 1/A, 2400 Dunaujváros, Hungary, [kovari@uniduna.hu](mailto:kovari@uniduna.hu)

<sup>5</sup> Kandó Kálmán Faculty of Electrical Engineering, Óbuda University, Bécsi street 96/B, 1034 Budapest, Hungary, [kovari.attila@uni-obuda.hu](mailto:kovari.attila@uni-obuda.hu)

<sup>6</sup> GAMF Faculty of Engineering and Computer Science, John von Neumann University, Izsáki u. 10, 6000 Kecskemét, Hungary [kovari.attila@gamf.uni-neumann.hu](mailto:kovari.attila@gamf.uni-neumann.hu)

**Abstract:** Natural Language Processing (NLP) is becoming increasingly important for researchers doing research in the field of information technology. However, NLP itself is not only used in the academic world, but is also a topic of interest in the competitive world, as it offers many opportunities for automation. The aim of this paper is to analyze and summarize the articles published on NLP from several perspectives, with a particular focus on the development and practical application of the technology, as well as to explore possible future research opportunities, directions and its potential role in industry. Among the practical applications of NLP, research and development in the field of education is increasingly being observed. In order to identify future uses, the underlying textual stocks of the technology should be considered. From this point of view, we can focus on the insurance and banking sectors, where there is a series of administrative processes where centralized paper processing is the incoming data set for the process and where a database dating back several years may be available. The research has confirmed that there are still many research opportunities in the field of NLP, as even the most modern methods cannot be applied without error and full context coverage. Moreover, the field of application can be extended considerably, both from the technical and from the economic and educational point of view.

**Keywords:** *natural language processing; deep learning; artificial intelligence;*

## Géppel történő természetes nyelvfeldolgozás üzleti alkalmazása

Krutilla Zsolt<sup>1,2</sup>, Kővári Attila<sup>3,4,5,6</sup>

<sup>1</sup> Szoftverfejlesztési és Alkalmazási Tanszék, Informatikai Intézet, Dunaujvárosi Egyetem, Táncsics Mihály út 1/A, Dunaujváros 2401, Magyarország, [krutillazs@uniduna.hu](mailto:krutillazs@uniduna.hu)

<sup>2</sup> *Alkalmazott Informatikai és Alkalmazott Matematikai Doktori Iskola, Óbudai Egyetem, Bécsi út 96/B, Budapest 1034, Magyarország*

<sup>3</sup> *Digitális Technológia Intézet, Informatikai Kar, Eszterházy Károly Katolikus Egyetem, Eszterházy tér 1, 3300 Eger, Magyarország kovari.attila@uni-eszterhazy.hu*

<sup>4</sup> *Műszaki Intézet, Dunaujvárosi Egyetem, Táncsics M. u. 1/A, 2400 Dunaujváros, Magyarország kovari@uniduna.hu*

<sup>5</sup> *Kandó Kálmán Villamosmérnöki Kar, Óbudai Egyetem, Bécsi út 96/B, 1034 Budapest, Magyarország kovari.attila@uni-obuda.hu*

<sup>6</sup> *GAMF Műszaki és Informatikai Kar, Neumann János Egyetem, Izsáki u. 10, 6000 Kecskemét, Magyarország kovari.attila@gamf.uni-neumann.hu*

---

**Absztrakt:** A számítógéppel történő természetes nyelvi feldolgozás, azaz az NLP (Natural Language Processing) egyre jelentősebb szerephez jut. Jelen tanulmány célja az NLP témában megjelent cikkek elemzése és több szempontú összefoglalása, különös tekintettel a technológia fejlesztésére és gyakorlati alkalmazására, valamint esetleges jövőbeli kutatási lehetőségek, irányvonalak feltárására és az iparban betöltött lehetséges szerepére. A jövőbeli felhasználási terület meghatározásához figyelembe kell vennünk a technológia alapjául szolgáló szöveges állományokat. Ezen aspektusból előtérbe helyezhetjük a biztosítási és banki szektorokat is, melyek esetében olyan ügyintézési folyamatok sorozata zajlik le, ahol központi papírfeldolgozás képezi a folyamat bejövő adatállományát és több évre visszamenőleges adatbázis állhat a rendelkezésünkre. A kutatás során beigazolódott az, hogy az NLP területén még számos kutatási lehetőségek előtt állunk, ugyanis a jelenleg legmodernebb módszerek sem alkalmazhatók hiba nélkül és teljes kontextus-lefedettséggel. Ezen felül az alkalmazási terület is jelentősen kiterjeszhető, mint a műszaki, mint a gazdasági és oktatási terület aspektusából.

**Kulcsszavak:** *természetes nyelvi feldolgozás; mélygépi tanulás; mesterséges intelligencia;*

---

## 1. Bevezető

Jelen cikk célja a természetes nyelv számítógépes feldolgozásának jelenlegi és jövőbeni alkalmazási területek lehetőségeinek feltárása, valamint az alkalmazott módszerekhez kapcsolódóan megjelent tudományos cikkek elemzése. Napjainkban egyre nagyobb teret hódít az informatikában a számítógépes természetes nyelvi feldolgozás, azaz az NLP (Natural Language Processing). Az NLP mint technológia ötlete azonban nem a 21. század. században fogalmazták meg, de már az 1950-es években is folytak releváns kutatások és születtek elméletek arra vonatkozóan, hogy egy gép hogyan tudja értelmezni az emberi beszédet. Az 1960-as években King, Masterman, Ceccato és Yngve olyan fordítógépen dolgoztak, amely magát az NLP technológiát alapozta meg. (Wilks, 2005) Az IT megjelenésével a RENDEZVOUS rendszer volt az első természetes nyelvi interfész (NLI) adatbázis. Az 1970-es években két transzformációs NLP megoldási módszer volt elterjedtebb, a generatív és a statisztikai algoritmus alapú módszer. A generatív eljárás esetében a két nyelv szerkezetét, szabályrendszerét veszi alapul, és meghatározza a két szabályrendszer közötti kapcsolatot. A statisztikai módszer esetében a tanítási minta valószínűségi szabályai alkotják az összefüggést.

(Kovács, 2009) Az NLP eredete tehát az ötvenes évekre vezethető vissza, jóval a számítógépek megjelenése előtt. Manapság elsősorban a Deep Learning eljárásokat alkalmazzák az NLP során, tekintettel arra, hogy ma már igen nagy pontosságú neurális hálózati modellek, valamint hatalmas adatbázisok és adattárházak állnak rendelkezésünkre a modellek betanításához, de korábban inkább a statisztikai és különféle matematikai modellek használata volt jellemző. A jelenlegi nyelvi modelleknek azonban megvannak a maga hátrányai is, különös tekintettel arra, hogy bár nagy adatbázis áll rendelkezésünkre a modellek tanítására, ezek korán sem nevezhetők tiszta adatoknak, sok esetben gépi generált tanulói adatokat használnak, amelyek torzítják az eredményt a „valódi beszéd” aspektusából. Végző célunk, hogy hosszú távon megtaláljuk azt az NLP technológiát, amely az emberhez hasonlóan képes elemezni és értelmezni a természetes beszédet.

Az NLP alatt olyan számítógépes megoldásokat értünk, amelyek képesek a mindennapi természetes szövegek nyelvi elemzésére, annak érdekében, hogy számos feladathoz vagy alkalmazáshoz emberszerű nyelvi feldolgozást érjenek el. A meghatározás azonban korántsem tökéletes, több megjegyzést is fűzhet hozzá. Mindenekelőtt nézzük meg közelebbről a „számítógépes megoldásokat”, mert számos módszer vagy technika közül választhatunk egy-egy adott típusú nyelvi elemzés elvégzéséhez. Másodszor a „természetes szövegek” jelentését is érdemes kiemelni, mivel ez azt jelenti, hogy a szövegek bármilyen nyelven, műfajban, írott vagy beszélt szövegek lehetnek. Az egyetlen követelmény az, hogy olyan nyelven legyenek, amellyel az emberek kommunikálnak egymással.

Fogalmilag mi emberek hajlamosak vagyunk nyelvi elemzési szinteket használni, mert mindegyik szint más-más jelentéstípust közvetít. Az NLP-rendszerek a nyelvi elemzés különböző szintjeit vagy szintek kombinációit használják, és ez a különböző NLP-alkalmazások közötti különbségekben is megmutatkozik. Ez sok zavart okoz a nem szakértők körében, hogy mi is az NLP valójában, mivel egy olyan rendszer, amely bármilyen Ezen elemzési szintek részhalmaza NLP-alapú rendszernek mondható, tehát a különbség valójában az lehet, hogy a rendszer „gyenge” vagy „erős” NLP-t használ. (Liddy, 2001)

## 2. Módszertan

Az elemzés célja, hogy az NLP témájához kapcsolódóan megjelent tudományos cikkeket tanulmányozzuk. Ehhez meg kell határozni az adott témát és ki kell tűzni a célokat. A kutatás lényegi része az NLP tudományág történetének és fejlődési szakaszainak, valamint a jelenlegi

alkalmazási területeknek az áttekintése. Ennek megfelelően került kiválogatásra a hazai és elsősorban a nemzetközi szakirodalom.

### 2.1. *Kutatás módszertana*

A cikkek kiválasztásához és az irány meghatározásához szisztematikusan a klasszikus kutatási módszertan alkalmazásával, az empirikus tapasztalatok és a kvalitatív elemzések eredményeinek ötvözése került alkalmazásra. A tudományági sajátosságok szempontjából a tárgyalt téma külön témakörökre bontható, így az alkalmazott módszerek témája, a tudományág eredete és fejlődése (valamint szakaszai), a nyelvészet és az informatika kapcsolata, mély tanulási alkalmazások fejlesztése, a matematikai nyelvi modellek fejlesztése, valamint a tanítási adatok struktúrái és tárolási technológiai kategóriákra bontható. Ezeknek a témáknak a közös elemzése azonban évekig is eltarthat, így ennek megfelelően a hatókör a tudományág fejlesztésére és az alkalmazott technológiák körére került redukálásra.

A szakirodalom áttanulmányozása során a keresők nemcsak releváns tudományos cikkeket, könyveket hoztak fel, hanem egyéb a témakörhöz köthető tanulmányokat is, melyek azonban nem relevánsak a jelen témakör aspektusából. A kutatási terület nagysága és aktualitása miatt az eredményeket szükséges volt leszűkíteni a releváns információkat tartalmazó eredményekre, amelyek elsősorban az NLP fejlesztésére, kezdetleges algoritmusaira és azokra a területekre koncentráltak, melyek egészen a deep learning (gépi mélytanulás) megoldásokig vezettek.

## 3. NLP módszerek és fejlődési szakaszai

Az NLP technológia terén nem csak általánosságban vehető szavakról és mondatokról beszélhetünk, hanem az úgynevezett stopszavakat is érdemes figyelembe vennünk. A stopszavak olyan csekély jelentéssel bíró szavak, amelyeket eltávolítanak az indexekből és a lekérdezésekből. Az NLP-ben a stopszavak negatív mintáknak is tekinthetők, amelyeket az LSTM neurális hálózatok tanításakor használnak. A stopszavaknak kevés jelentése lehet a gyakoriság vagy a fogalom szempontjából, továbbá a nagyon gyakori szavak eltávolítása nem befolyásolja a dokumentumok rangsorolását és jelentését sem. Ha a szavaknak kevés fogalmi jelentésük van, akkor eltávolíthatók, függetlenül attól, hogy gyakoriságuk magas vagy alacsony a gyűjteményben. Valójában különösen akkor fontos ezeket a szavakat eltávolítani, amikor alacsony a gyakoriságuk, mert ezek a szavak befolyásolják leginkább a dokumentumok rangsorolását. Szövegelemzés tekintetében elmondható, hogy szakirodalom általában kétféle szövegelemzési technikát tárgyal, a kulcsszószámlálást és a térképelemzést. Ezeken felül (vagy

ezzel egyidőben) dokumentumindexelésről is beszélhetünk, melynek során a lekérdezés feldolgozás tipikus megközelítése a következő. Először egy tokenizálási folyamat megy végbe, majd a stopszavak eltávolításra kerülnek. Ezen túlmenően a természetes nyelvi feldolgozási technikák azonosíthatják a kifejezéseket vagy az osztott összetételeket. (Hiemstra és Jong, 2001)

Az első módszer, a kulcsszószámlálás, jellemzően számítógép segítségével méri automatikusan azon kulcsszavak gyakoriságát, amelyekről a kutatók korábban megállapították, hogy egy érdekes konstrukcióhoz kapcsolódnak (a mai napig ez a módszer volt a legszélesebb körben használt technika). (Duriau, Reger, Pfarrer, 2007) A technika népszerűsége abból adódik, hogy a kulcsszavak gyakorisága objektív és intuitív mutatója az gyűjtemény méretének vagy relatív fontosságának. (Knoke, 1982)

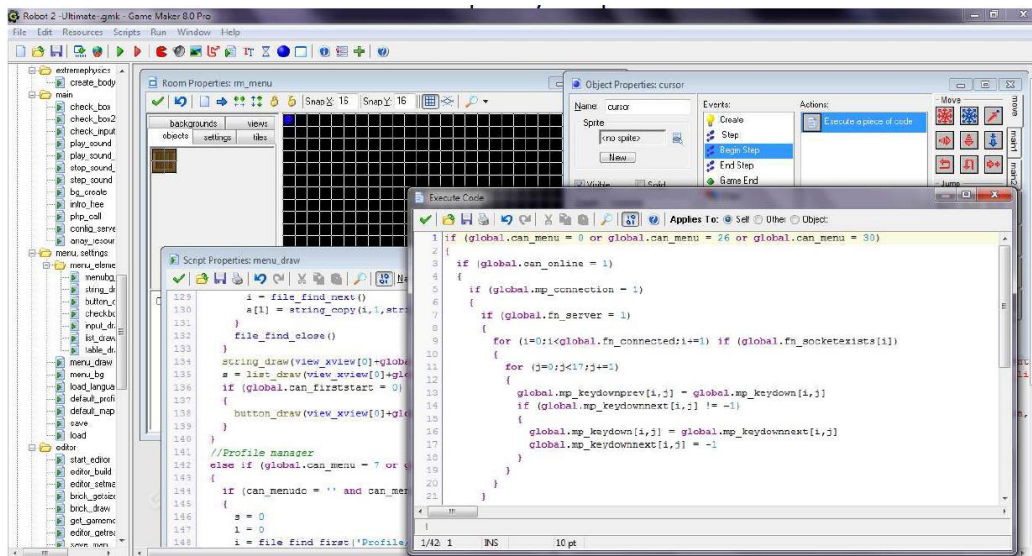
A második módszer, az úgynevezett „térképezési elemzés”, melynek során a fogalmak közötti relációs kapcsolatok jelennek meg a szövegekben (többnyire ez inkább az fenti módszert kiegészítő technika). (Axelrod, 1976) A szövegelemzőket általában arra képezik ki, hogy azonosítsák a releváns fogalmakat és összefüggéseket a szövegekben, és addig olvassák át a dokumentumokat, amíg el nem jutnak a szövegek mögött meghúzódó kognitív kapcsolatok megbízható reprezentációjához. (Huff, Narapareddy, Fletcher, 1990)

Bár mindkét technika elfogadott szövegelemzési módszerré vált a szakirodalomban, a két hagyományos technika nyilvánvaló kompromisszumot jelent. A kulcsszavak megszámlálásával a kutatók szűk mértékegységeket vonhatnak ki a szöveg tartalmából, de nem tudják kihasználni azt a részletes és értékes információt, amely a szövegek szavainak tágabb szerkezetében rejlik. (Carley, Kathleen, Palmquist, 1992) A gépi segítséggel támogatott NLP módszerek legújabb fejlesztései új lehetőségeket kínálnak a szöveges adatok elemzésére és értelmezésére. (Grimmer, 2013) A modern NLP lehetővé teszi a kutatóknak, hogy számítási algoritmusok segítségével mélyebb jelentésstruktúrákat vonjanak ki nagy mennyiségű szövegből, és ennek segítségével megnyílik az elemzés lehetősége.

Összefoglalva elmondható, hogy a szövegbányászat és a természetes nyelvi feldolgozás a „természetes nyelv” megértését és elemzését jelenti számítógépes algoritmusok és programok segítségével, és fontos kutatási irány a mesterséges intelligencia alkalmazási területén. Folyamatos és kiterjedt gépi kutatásokkal. tanulási és adatbányászati algoritmusok, a meglévő szövegbányászati technológiák jó eredményeket értek el az automatikus absztrakcióban, az automatikus kérdésválaszolásban, a webes relációs hálózatelemzésben és az anafora felbontásban (Zeng et al, 2015).

#### 4. Az NLP szerepe a gyakorlatban

Az NLP általában olyan oktatási szoftverrendszerek és oktatási stratégiák fejlesztésére összpontosít, amelyek támogathatják a természetes nyelvek használatát az oktatásban (pl. e-Assessment és Text Adapter). Az NLP-vel rendelkező szoftverrendszerek természetes körülmények között képesek azonosítani a nyelvtanulás folyamatát. (Gy. Molnár; P. Nyíró, 2016)



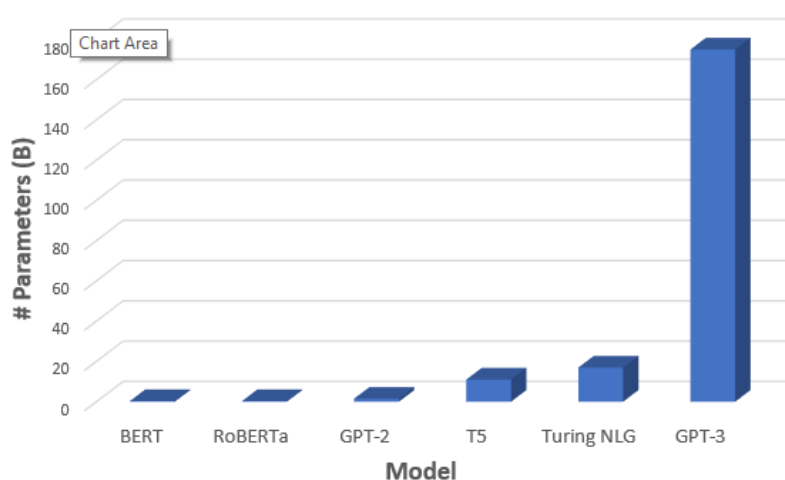
1. ábra - Játékfejlesztés a kurzusokon alkalmazott Game Maker 8.0 használatával. (Gy. Molnár; P. Nyíró, 2016)

Az NLP hatékony módszer egy olyan rendszer kifejlesztésére is, amely különféle szavakon, mondatokon és szövegeken keresztül dolgozza fel a nyelvi bemenetet a természetes környezetben. A természetes nyelvi feldolgozás különféle nyelvtani szabályokat és nyelvi megközelítéseket is használ, mint például a származékokat, a nyelvtani igeidőket, a szemantikai rendszert, a lexikont, a korpust, a morfémákat, az igeidőket stb.

Ezek a hatékony megközelítések felhasználhatók az oktatási környezetben is, hogy a tanulók jobban megértsék az oktatási anyagokat. Az NLP a nyelvtanulás területén is széles körben ismert megoldás az egész világon, és sikeresen alkalmazzák a nyelvoktatási rendszerek fejlesztésének hatékony eszközeként. Hasonlóan hatékony és eredményes tanulás érhető el vizuális alapú, problémamegoldó, algoritmikus gondolkodással (Francisti et al, 2021). A legtöbb tanulmányban az angol a leggyakoribb nyelv, ami azt mutatja, hogy az NLP-t hatékonyan használják a nyelvtanulásban. Az NLP hatékony megközelítés az arab országok oktatási rendszerének javítására is. (Habash, 2010)

Az elmúlt évtizedben a mesterséges intelligencia területén a technikai fejlődés jelentős áttörést hozott az AI-technológia kutatásával és alkalmazásával foglalkozó valamennyi területen, így az NLP területén is. Az első, mondhatni kézzelfogható áttörésre a számítógépek megjelenéséig kellett várni, amelyek képesek voltak elméleteket a gyakorlatba átültetni és sikerüket bizonyítani.

A Moore-törvénynek köszönhetően a számítási teljesítmény évtizedek óta tartó exponenciális növekedése lehetővé tette az olyan statisztikai megközelítéseket, mint a GPT-3. Ezek a modellek az írott nyelv tömeges példáival való találkozás révén tanulják meg a nyelvet. (ELKINS, Katherine; CHUN, Jon, 2020)



2. ábra - GPT-3: The New Mighty Language Model from OpenAI (forrás: <https://towardsdatascience.com/gpt-3-the-new-mighty-language-model-from-openai-a74ff35346fc>)

A GPT-3 (Generative Pre-trained Transformer) egy harmadik generációs, autoregresszív nyelvi modell, amely mély tanulást használ az emberhez hasonló szöveg előállítására. Vagy egyszerűbben fogalmazva, ez egy olyan számítási rendszer, amelyet arra terveztek, hogy szavak, kódok vagy más adatok szekvenciáit generálja egy forrásbemenetből, az úgynevezett promptból kiindulva. A GPT első, 2018-as iterációja 110 millió tanulási paramétert használt. Egy évvel később a GPT-2 már 1,5 milliárdot használt belőlük. Azonban mindezen számok eltörpülnek a jelenlegi GPT-3 175 milliárd paraméteréhez képest. A modell tanítása a Microsoft Azure mesterséges intelligencia szuperszámítógépén (Scott 2020) történik. A modell felhasználási területe igen széles skálán mozog, beleértve az összegzést, a fordítást, a nyelvtani javítást, a kérdések megválaszolását, a chatbotokat, az e-mailek összeállítását és még sok más is. (Floridi, Chiriatti, 2020)

## 5. Következtetés, további kutatási lehetőségek

A kutatás során beigazolódott az az elképzelés, hogy a tudományág korai szakaszában és a számítógépek megjelenése előtt a matematikai modellezés és a statisztikai valószínűségekre alapuló modellek teret hódítottak a területen. Az információs technológia (számítógép) megjelenésével a módszerek és az alkalmazott eljárások is megváltoztak, úgymond követték a technológiai fejlődést, és egyre inkább hasznosították a technológiában rejlő erőforrásokat. A számítástechnika folyamatos fejlődésével megjelentek az adatbázisok, amelyek új alapokra helyezték az NLP tudományát. Az adatvezérelt szövegelemzés korszaka kezdődött el, amikor lehetővé vált nagy adathalmazok tárolása, kezelése és elemzése. Új eszköz állt a kutatók rendelkezésére, amely még nem nevezhető mesterséges intelligenciának, de már a logikai megközelítés sokkal inkább hasonlított a mai LSTM (Long-Short-Term-Memory) neurális modellhez, mint a statisztikai elemzéshez. Lehetőség nyílt a korábbi adatok alapján történő szövegelemzésre, megfigyelésre és azonosításra. Az NLP tudományágában ez a korszak tekinthető az egyik legjelentősebb szakasznak, mondhatni paradigmaváltásnak.

Bár a korábbi eredmények alapján az NLP mint tudományág még az adatvezérelt módszer korszakában is jelentős eredményeket ért el, de az igazi áttörést a mesterséges intelligencia megjelenése és sikeres megvalósítása hozta meg. Az újonnan megjelenő mély tanulási módszer új perspektívába helyezte az NLP-kutatással foglalkozó tudósok munkáját, új megközelítések és módszerek jelentek meg. A tanulmány során betekintést nyertünk a természetes nyelv gépi feldolgozásának kiindulópontjába, feltárta, hogy a diszciplína elmélete jóval az első számítógépek megjelenése előtt született, és annak ellenére, hogy egyre nagyobb áttöréseket érnek el a deep learning módszer segítségével, még mindig nem áll a rendelkezésünkre egy olyan eljárás vagy módszertan, amelyet valóban az NLP probléma megoldására terveztek. Az NLP a mai napig tudományosan megoldandó probléma, amelyre még mindig nincs pontos és egységes megoldás.

A GPT-3 modellek megjelenésével egy sokkal hatékonyabb nyelvi elemző eszköz áll a rendelkezésünkre és a fordítógépek mellett egyre nagyobb az igény a természetes nyelvi elemzésre is, amely nemcsak az emberi nyelven írt szövegek gépi értelmezésére ad lehetőséget, hanem egy olyan kommunikációs eszközt is biztosít számunkra, amely új szintre emelheti a számítógépes kommunikációt (J. Katona, 2021a), (J. Katona, 2021b), (J. Katona, 2022). Képzelnünk el egy olyan programozási feladatot, melyet már nem emberek programoznak, hanem „természetes beszéd útján felmondjuk” a szoftverrel szemben támasztott elvárásainkat a



számítógépnek, melyet a gép képes értelmezni, majd a fejlesztést önállóan elvégezni a szóbeli specifikációnak megfelelően. Természetesen a természetes nyelv ilyen mélységű felhasználásáig több problémát is meg kell oldanunk, azonban az ilyen mélységű természetes beszéd értelmezése már egy másik szintet képvisel, az értelem szintjét.

Egy másik (és természetesen sokkal közelebbi) felhasználási területe lehet az NLP-nek a banki és biztosítási dokumentumok értelmezése és a további munkavégzés (feladatlépés) elvégzése. Jelenleg ezen a két területen igen nagy jelentőséggel bírnak a szöveges dokumentumok értelmezése, így az NLP egy új szintre emelheti azirodai munkavégzés folyamatát.

Az NLP aspektusából ugyancsak izgalmas terület lehet a jogrendszerek labirintusában történő kiigazodás és jogesetek azonosítása akár a magánszférában, akár a bírósági döntéshozás területén. A jogi szabályozás tekintetében az NLP technológia által a döntéshozás akár automatizálható folyamattá válhat, azonban semmiképp sem hanyagolhatjuk el a területen esetlegesen betölthető jelentőségüket, szerepüket (Riczu, Krutilla, 2021).

### **Irodalomjegyzék**

Axelrod, R. (1976). *The cognitive maps of political elites. Structure of decision.* Princeton University Press: Princeton, NJ.

Carley, K., & Palmquist, M. (1992). Extracting, representing, and analyzing mental models. *Social forces*, 70(3), 601-636.

D. a. JHK Knoke (1982). *Network analysis*, Beverly Hills: SAGE Publishing.

Duriau, V. J., Reger, R. K., & Pfarrer, M. D. (2007). A content analysis of the content analysis literature in organization studies: Research themes, data sources, and methodological refinements. *Organizational research methods*, 10(1), 5-34.

Elkins, K., & Chun, J. (2020). Can GPT-3 pass a Writer's turing test?. *Journal of Cultural Analytics*, 5(2), 17212.

Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681-694.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political analysis*, 21(3), 267-297.

Habash, N. Y. (2010). Introduction to Arabic natural language processing. *Synthesis lectures on human language technologies*, 3(1), 1-187.

Hiemstra, D., & Jong, F. D. (2001). *Statistical Language Models and Information Retrieval: natural language processing really meets retrieval.*

Huff, A. S., Narapareddy, V., & Fletcher, K. E. (1990). Coding the causal association of concepts. *Mapping strategic thought*, 311-325.

Katona, J. (2021a). Clean and dirty code comprehension by eye-tracking based evaluation using GP3 eye tracker. *Acta Polytechnica Hungarica*, 18(1), 79-99.

Katona, J. (2021b). Analyse the Readability of LINQ Code using an Eye-Tracking-based Evaluation. *Acta Polytech. Hung*, 18, 193-215.

Katona, J. (2022). Measuring Cognition Load Using Eye-Tracking Parameters Based on Algorithm Description Tools. *Sensors*, 22(3), 912.

Kovács, L. (2009). Szemantikai gráf alapú mondatelemző modul kidolgozása IS-NLI értelmezőhöz, VI. Magyar Számítógépes Nyelvészeti Konferencia, pp. 356-359.

Liddy, E. D. (2001). *Natural language processing*.

Molnár, Gy et al (2016). A gyakorlati programozás tanításának játékfejlesztésen alapuló, élménypedagógiai alapú módszerének bemutatása. Karlovitz János Tibor (szerk.), *Pedagógiai és szakmódszertani tanulmányok*, 89-98.

Riczu, Z., & Krutilla, Z. (2021). The impact of optical character recognition artificial intelligence on the labour market. *International Journal of Engineering and Management Sciences*, 6(4).

Wilks, Y. (2005). The history of natural language processing and machine translation. *Encyclopedia of Language and Linguistics*, 9.

Zeng, Z., Shi, H., Wu, Y., & Hong, Z. (2015). Survey of natural language processing techniques in bioinformatics. *Computational and mathematical methods in medicine*, 2015.

### **Rövid szakmai életrajz**

**Krutilla Zsolt** a Dunaújvárosi Egyetemen egyetemi tanársegédként dolgozik az Informatika Intézetnél. 2007 óta dolgozik az OTP Bank Nyrt.-nél, 2021 óta pedig mint főosztályvezető. Dunaújvárosi Egyetemen szerzett mérnökinformatikus (szoftvertechnológia) és okleveles mérnöktanár (informatika) diplomát. Kutatási területe a mesterséges neurális hálózat alapú, Magyar nyelvű természetes-nyelvi (NLP/NLU) és a szótáralapú szöveganalitikai modellek vizsgálata banki back-office környezetben.

**Kővári Attila** az Eszterházy Károly Katolikus Egyetem egyetemi tanára. Informatika és neveléstudományok területén habilitált. Kutatási területe az alkalmazott informatika, adatfeldolgozás, adatelemzés valamint a digitális pedagógia, IKT alkalmazása az oktatásban és kapcsolódó témaköröket öleli fel.