conTEXT 2023

Change the Game?

CEU, 2023.11.14.

ISBN 978-615-01-9105-8

A conTEXT 2023 konferencia szervezője:

Clementine/Statistical Products Hungary Kft.

1115 Budapest, Bartók Béla út 105-113. 1/b.

A konferencia házigazdája:

Körmendi György Olivér

Clementine

ügyvezető igazgató

Szerkesztette[1]:

Izsán Orsolya, Keresztesi Ildikó, Pancza Judit

[1]oktatas@clementine.hu

Tartalomjegyzék

Mérföldkövek a ChatGPT-ig

Yang Zijian Győző

HUN-REN Nyelvtudományi Kutatóközpont
yang.zijian.gyozo@nytud.hun-ren.hu

Kulcsszavak: nyelvtechnológia, transzformer, nagy nyelvi modellek, ChatGPT

Az elmúlt időszak legnépszerűbb témája volt a ChatGPT sikere. Rendkívül nagy hatása lett a hétköznapjainkban, a vállalatok körében és a tudomány világában is. Új piaci szereplők jöttek létre, akik eddig nem foglalkoztak nyelvtechnológiával. A sajtónak köszönhetően mindennapos témává vált a mesterséges intelligencia és ez a fogalom összefonódott a ChatGPT-vel és a hozzá hasonló alkalmazásokkal.

Nagy utat járt be a nyelvtechnológia tudományága. A szabályalapú megoldásoktól a statisztikai módszereken át a neurális modellekig. Az igazi áttörést a neurális módszerek hozták, azon belül is a transzformer architektúra. A transzformer segítségével a modell rendkívül finom környezetfüggő összefüggéseket és tudást képes megtanulni a szövegből. Megjelenése átalakította a folyamatokat a nyelvtechnológiában, a feladatmegoldás kétlépcsős lett: előtanítás és finomhangolás. Az előtanítás során egy nyelvmodellt tanítunk általános nyelvtudásra, majd a finomhangolás útján tovább tanítjuk az előtanított nyelvmodellt feladatspecifikus tudásra. Ezzel a megoldással szinte minden nyelvtechnológiai feladatban piacvezető (*state-of-the-art*) eredményeket értek el.

A transzformer modellben rejlő potenciált megsejtették a nagy vállalatok és belekezdtek egy nyelvmodellezési versenybe. Az elmúlt években nagyobbnál nagyobb nyelvmodelleket tanítottak nagyobbnál nagyobb adathalmazokon. Jelenleg az élen az OpenAI jár a GPT (*Generative pre-trained transformers*) modellekkel és az azok finomhangolásával előállított ChatGPT alkalmazással. Jelenleg az egyik legnagyobb nyelvmodell ami létezik az a GPT-4, ami 1,76 billió paraméterszámmal rendelkezik. A GPT-4 egy multimodális modell, ami a szöveg mellett képet is képes feldolgozni. A ChatGPT sikerének egyik kulcsa, hogy egy ilyen nagy nyelvi modell (*large language model – LLM*) áll mögötte és szolgáltatja a tudásának az alapját. Másik kulcsfontosságú rész az emberi visszajelzések integrálása. A modell a finomhangolás során a megerősítéses tanulás módszerével figyelembe veszi az emberi visszajelzéseket. Ezzel a módszerrel a modell megtanul sokkal „emberibb” válaszokat adni.

A ChatGPT sikere új utakat nyitott meg, a kezdeti fellángolás után megmutatkoztak a problémák és hiányosságok is, mint a hallucináció, a tények pontatlan tudása, vagy az erőforrás igénye egy ilyen nagy nyelvi modellnek. A problémák kezelésére számos kutatás és megoldás jelent már meg, de még mindig van fejlődési tere. Végül, de nem utolsó sorban ne feledkezzünk meg a régebbi hagyományosabb módszerekről, amelyek például erőforrás hiányában továbbra is tökéletesen meg tudnak oldani egyszerű nyelvtechnológiai feladatokat.

Engedjük szabadjára a WatsonX-et

Huszti Dániel

IBM Magyarországi Kft., Budapest, Magyarország, daniel.huszti1@ibm.com

Kulcsszavak: mesterséges intelligencia, generatív MI, nagy nyelvi modell, LLM, RAG, élő bemutató

A ChatGPT megjelenése óta mindenki a nagy nyelvi modellek és a generatív mesterséges intelligencia (MI) által előidézett technológia forradalomról beszél, azonban a legtöbb vállalat bizonytalan milyen területen és miként használja ki ezen új MI technológia adta lehetőségeket. Ennek főbb okai a technológiába vetett bizalom hiánya, mi történik, ha tévesen válaszol, a szigorú törvényi és adatvédelmi szabályozásoknak való megfelelés, illetve az új technológia bevezetésével járó megtérülési költségek kalkulálhatóságának nehézsége.

Új megközelítésben a generatív MI

Az IBM májusban bejelentett megoldása, a Watsonx adat és mesterséges intelligencia platformja új alapokra helyezi a generatív MI felhasználását. Kifejezetten nagyvállalati igényekre lett kialakítva, azonban rendelkezik a modellek által hozott döntések felügyelhetőségével is, így növeli a bizalmat az MI technológiával szemben.

Az általános célú nagy nyelvi modellek, mint például a ChatGPT helyett több kisebb célorientált modell alkalmazásával és finomhangolásával megbízhatóbb, kevesebbet tévesztő és jelentősen olcsóbb eredmények érhetők el. Nyitott platform révén az IBM által fejlesztett iparágspecifikus modelleken felül külső gyártói és IBM által bevizsgált nyílt forráskódú modellek egyaránt alkalmazhatók, saját adathalmazzal személyre szabhatók vagy akár saját modellek is fejleszthetők az IBM új szuperszámítógépén.

A Watsonx futtatókörnyezete lehet akár felhős szolgáltatás alapú (SaaS) vagy lokálisan telepíthető is, így akár a legszigorúbb előírásokkal rendelkező pénzügyi és államigazgatási ügyfelek számára is lehetőséget biztosít ezen új technológia felhasználására.

Felhasználási esettanulmányok és élő bemutató

A generatív MI alkalmazási esetei sokrétűek, a belső tudástár feldolgozásától kezdve, az ügyfélszolgálati esettanulmányokon, email- és tartalomgeneráláson, összefoglaláson, fordításon, kódgeneráláson át, dokumentum vagy internetes tartalom feldolgozásra egyaránt hatékonyan alkalmazható.

Az előadás során pár esettanulmány részletes kifejtésén felül sor kerül két magyar nyelvű élő bemutatóra is. Az elsőben egy PDF dokumentum feltöltését, feldolgozását és az arra feltett kérdések megválaszolását ismertetjük, kitérve miként lehet elősegíteni a bizalmat a technológia által generált válaszokban. A második bemutató azt mutatja be, hogyan lehet integrálni az IBM Watsonx Assistant virtuális asszisztensét a generatív MI technológiával, ezzel javítva az ügyfélszolgálati virtuális asszisztens hatékonyságát, eközben megőrizni a költséghatékonyságot és a döntési kontrollt az MI felett. Az élő bemutató elkészítése során az IBM Watsonx.ai, Watsonx Assistant szolgáltatásokat, Langchain Python könyvtárat és vektoradatbáziskezelőt alkalmaztunk.

Szóbeágyazások és nagy nyelvmodellek társadalomtudományi alkalmazásának példái

Rakovics Zsófia[1, 2]

[1] Eötvös Loránd Tudományegyetem, Társadalomtudományi Kar, Társadalomkutatások Módszertana Tanszék, Research Center for Computational Social Science, Szociológia Doktori Iskola, Budapest, Magyarország, zsofia.rakovics@tatk.elte.hu
[2] HUN-REN Társadalomtudományi Kutatóközpont, CSS-RECENS, MTA–TK Lendület Társadalmi Rétegződéssel Foglalkozó Digitális Társadalomtudományi Kutatócsoport, Budapest, Magyarország, rakovics.zsofia@tk.hu

Kulcsszavak: számítógépes társadalomtudomány, természetesnyelv-feldolgozás, időbeli szóbeágyazás, nagy nyelvmodell, chatGPT

Módszertani szemléletű előadásomban a szóbeágyazási eljárások és a nagy nyelvmodellek társadalomtudományi alkalmazásának egy-egy példáját mutatom be, reflektálva a természetesnyelv-feldolgozás és a számítógépes társadalomtudomány használatának elterjedésére. Az elemzésekhez használt adatok politikusi beszédek leiratai, illetve a nagy nyelvmodellek által generált szöveges kimenetek.

A szóbeágyazási vektortérmodellek segítségével a szavak egy olyan numerikus reprezentációját határozhatjuk meg, amely kódolja azok szövegben elfoglalt helyét és más szavakhoz való viszonyát. Az általánosan használt szóbeágyazási eljárások statikusak (word2vec, PPMI), nem veszik figyelembe az idő dimenzióját, ezért ezek nem alkalmasak a szavak szemantikai modellezésére. Amennyiben a szavak jelentésváltozását szeretnénk vizsgálni, egyszerű megoldásként adódhat, hogy külön beágyazásokat illesztünk az adatok különböző időegységeire, de ehhez minden egység esetén kellően nagy alkorpuszra van szükségünk, ami sokszor nem áll rendelkezésünkre. A legújabb modellek ezért időegységek közötti információmegosztást alkalmaznak a beágyazások adattakarékosabb illesztéséhez, alapozva arra a megfigyelésre, mely szerint a legtöbb szó csak kis jelentésváltozáson megy keresztül rövid idő alatt. Előadásomban bemutatom a statikus PPMI (Positive Pointwise Mutual Information) szóbeágyazási eljárás egy módosított változatát, amely lehetővé teszi a szavak időbeli szóbeágyazását: A modell első lépésben az egyes időszeletekre vonatkozóan kiszámolja a PPMI mátrixokat, melynek oszlopai ugyanazokat a kontextusszavakat tartalmazzák, biztosítva az időegységek beágyazási terének összehangolását. Második lépésként a modell a PPMI-mátrixok értékeit egy regularizált spline segítségével időben simítja, stabilizálva az egyes időegységek szóvektorait.

A mélytanuláson (deep learning) alapuló nagy nyelvmodellekre (large language model) épülő mesterséges intelligencia megjelenése új és eddig kiaknázatlan lehetőséget teremt a társadalomkutatási megismerés módszerei szempontjából. Reális lehetőséggé vált valós személyek – mint kvalitatív vagy kvantitatív adatforrások – helyett a nagy nyelvmodellek által szimulált fiktív, virtuális személyek bevonása adatközlőként, minden olyan kutatás esetében, amelynél a nyelv közvetíti az empíriát.

Előadásomban először az időbeli szóbeágyazási modell módszertani hátterét ismertetem és társadalomtudományi szempontból releváns kérdések kulcsszavainak jelentésváltozását mutatom be politikusi beszédek alapján, majd pedig ehhez kapcsolódó elemzési példákat hozok a chatGPT, mint egy ismert nagy nyelvmodell, használatával. Az eredmények összefoglalása mellett kitérek az egyes módszerekben rejlő lehetőségekre és korlátokra is.

A *következő* már az lesz! Fejezetek a szöveganalitikai modellek (végtelen) evolúciójából

Galántai László

OTP Bank Nyrt., Budapest, Magyarország, Laszlo.Galantai@otpbank.hu

Kulcsszavak: szöveganalitika, szövegbányászat, NLP, NLU

Az előadás célja, hogy a statisztikai algoritmusok üzleti felhasználását technológiatörténeti és szervezeti kontextusban tárgyalja. A szöveganalitikai modellek nagy utat jártak be a máig jól alkalmazható sql-szövegfüggvényektől a mostanában éppen alkalmazásba kerülő generatív nyelvmodellekig. Ezek, bár számítástudományi szempontból természetesen nagyon eltérőek, végsősoron ugyanarra az üzleti igényre adnak választ: a strukturálatlan szövegadatból automatikusan kinyerendő információ problémájára. Az előadás során röviden áttekintjük a szöveganalitikai evolúció főbb állomásait, az ontológiaialapú szakértői szótármodelleket, a gépi tanulásos és neurális hálós nyelvfeldogozókat és végül a generatív nagy nyelvi modelleket. Fő állításunk, hogy a forradalminak beállított mesterséges intelligencia inkább mesterséges, mint intelligens, és legfőbb eredménye, hogy képes strukturálatlan adatinput alapján automatikusan döntéseket hozni, vagyis egyfajta döntéstudományi, és nem kognitív tudományos intelligenciafogalom áll mögötte. Üzleti alkalmazása ezzel együtt számos és értékteremtő, az eredményeket azonban nem a szoftverek kognitív fejlődése (lévén ilyen nem volt), hanem a hardverek kapacitásbővülése és költségcsökkenése alapozta meg. A természetesnyelv-feldolgozás üzleti alkalmazásának tárgyalása során felülértékelt a szoftveres képességek bővülése, mert egyszerűen arról van szó, hogy korábban is ismert vagy azokra építetten fejlesztett algoritmusok vehetőek használatba a rendelkezésre álló, költséghatékonyan elérhető adattárolási és -feldolgozási kapacitások miatt, nem mellékesen ezek teszik lehetővé az üzleti folyamatok digitalizációját, vagyis számítógépes végzését és logolását, amely megnyitja az adatok rendelkezésre állásának lehetőségét. Kevés szó esik továbbá az alkalmazás szervezeti kontextusáról, a szolgáltatásokra irányuló dinamikusan bővülő keresletre válaszoló, a belső folyamatokat rendszeresen átalakító változásokról, a folyamatok permanens evolúciójáról, amit a gépi tanulás éppen csak követni tud.

Nyelvi modellek a programozásban

Schäffer Krisztián

aijunior.dev, Budapest, Magyarország, schaffer.krisztian@gmail.com

Kulcsszavak: munkamemória, kognitív terhelés, forráskód generálás

A nagy nyelvi modellek lenyűgöző képességeket mutatnak programozási feladatok megoldásában, kiemelkedő megértést tanúsítva mind a logika, mind a szintaxis terén.
Képesek például egy menetben működő kétdimenziós fizikai naprendszer szimuláció létrehozására webtechnológiák segítségével. A GPT-4 kiemelkedik a mezőnyből, de a konkurens modellek, bár még érezhető lemaradásban vannak, szédületes tempóval javulnak.
Jogosnak tűnik a feltételezés, hogy a fejlődés még nem tetőzött.
A modellek kognitív terhelhetősége már most lenyűgöző, képesek tucatnyi különböző követelményt és parancsot párhuzamosan menedzselni. Ez lehetővé teszi számukra a komplex kódolási projektek kezelését, különböző kódolási keretrendszerekhez való alkalmazkodást, és kevésbé gyakori programozási feladatok kezelését, melyeket részletesen meg kell fogalmazni a követelményben.
Ugyanakkor minél több követelményt támasztunk, például formázással vagy projekt-specifikus kódszervezéssel kapcsolatban, annál valószínűbb, hogy egyet vagy többet ezek közül a modell "elfelejt". Hogy pontosan melyiket, az részben véletlenszerű, részben a követelmény megfogalmazásának és promptban elfoglalt helyének függvénye: A fő task leírása után közvetlenül adott utasításokat például valószínűbb, hogy észben tartja.
Ebben nagyon hasonlóan viselkednek a modellek az emberhez, abban azonban már jelentős különbség van, hogy a modell szekvenciális jellege korlátozza az önellenőrzést: A feladat végeztével nem tud egyesével végigmenni a követelményeken és ellenőrizni őket, de ha erre ún. actor-critic architektúra használatával a kimenetet újabb promptokba visszacsatolva direkt megkérjük, előfordul, hogy észreveszi a hibát.
A munkamemória kérdése kritikus, mivel a jelenleg nagynak hívott nyelvi modellek erősen korlátozottak a bemeneti tokenek maximális számát illetően (ún. context window), működésük pedig állapotmentes - Minden új token generálásához az egész promptot és az addig generált kimenetet nulláról indulva teljesen feldolgozzák -, így az egyidejűleg feldolgozható karakterek száma csupán kilobyte nagyságrendű. A teljes fejlesztés alatt álló programkód nem fér bele a promptba, így központi kérdéssé válik, hogy a kód mely részei legyenek a középpontban.
Vektor-adatbázisok használatával a probléma csak részben oldható meg, hiszen egy-egy részfeladat kódolásához több, egymással nem kapcsolódó forrásfájlra is szükség lehet, melyeket a feladatleírás alapján nem biztos, hogy megtalálunk. Ilyen esetben a feladat dekompozíciója, vagy a manuális fájlkiválasztás segíthet. Ez utóbbi további előnye, hogy kontextust ad, egyszerűsítve a feladatleírást.
Továbbá, a modellek nem hibátlanok, és kudarcot vallhatnak, különösen, ha a promptokat nem megfelelően határozzuk meg. Például a modell felesleges kódrefaktorálásba kezdhet,
ha az utasítások nem világosak és pontosak, vagy egyes projekt-specifikus követelmények - melyek minden promptba bekerülnek - túl hangsúlyosak, miközben nem relevánsak az aktuális feladat szempontjából.
A jól meghatározott promptok kidolgozásának fontosságát nem lehet túlságosan hangsúlyozni - ez egy olyan készség, amelyet a fejlesztőknek el kell sajátítaniuk, hogy a nyelvi modellek teljes potenciálját hatékonyan kiaknázhassák. Igaz ugyan, hogy a tudás egy része

externalizálható például prompt template-ek segítségével, de ez növeli a modell kognitív terhelését, ami végső soron a teljesítmény romlásához vezethet.
Összefoglalva, a nyelvi modellek programozásban való felhasználásának potenciálja figyelemre méltó, és máris a fejlesztők hatékonyságának növekedését hozta, miközben további gyors fejlődés várható. A modellek kognitív teherviselő képessége és munkamemóriájuk korlátossága, valamint a prompt specifikációjának kihívásai jelentősek, de legyőzhetők készségfejlesztéssel és megfelelő eszközök használatával.

Szöveganalitikai technikák a gyakorlatban

Tamási-Mészáros Evelin

Clementine, Budapest, Magyarország, emeszaros@clementine.hu

Kulcsszavak: szöveganalitika, nyelvi modellek, nyílt forráskódú programozás, IVR, gépi intelligencia

A technológiai fejlődés ütemének gyorsulása a számítógépes nyelvészet területét is áthatja: a nyílt forráskódú programnyelvek térhódítása és a nyelvi modellek bővülése mind hozzájárul az alkalmazott szöveganalitikai technikák átalakulásához. Az adatelemzés és az informatika területe egyre szorosabb kapcsolatban áll egymással, amely a hagyományos elemzési eszköztár és módszertan átalakulásához vezet. A szakma is átalakulóban van, hiszen az autodidakta tanulás egyre nagyobb térhódításával nem csupán a szűk értelemben vett Data Scientistek vannak jelen a piacon, hanem ún. "Citizen Data Scientistek" [1] is demokratizálják a tudományterületet. A nyelvi modellek nagyságát nagyban befolyásolja a rendelkezésre álló írásos, illetve hangalapú adatok mennyisége és minősége, és ezért sok esetben a technológia óriások (pl. Google) bírnak potenciális előnnyel.

A szakma demokratizálódása azonban lehetővé teszi a kisebb cégek számára is, hogy olyan megoldásokat hozzanak létre, amelyek megkönnyíthetik az emberek mindennapi életét. A rendelkezésre álló nyelvi modellek egyik tipikus felhasználási területe az ügyfélszolgálat, hiszen itt nagyon gyakori és irányított a kommunikáció.

A továbbiakban a változások miként jének bemutatásaként olyan esettanulmányok kerülnek a középpontba, amelyek ember és ember közötti párbeszédekhez köthetőek, és mint gépi-intelligencia rendszerek hozzájárulnak az ügyfélszolgálatok robotizációjához. Egy ilyen megoldás hatékony működéséhez elengedhetetlen az adott szakterület és a szöveganalitika szoros együttműködése, továbbá hangalapú kommunikáció esetén a hanganalitika is előtérbe kerül.

A Clementine első megoldásaként Hangát lehet említeni: Hanga a párbeszédek bonyolultságát tekintve egy kevésbé összetett, csupán kérdés-válasz páron alakuló banki asszisztens. Működését tekintve képes az ügyfél kérését anélkül értelmezni, hogy az illetőnek el kellene navigálnia egy bonyolult telefonos menürendszerben – legyen a kérés akár egy egyenleglekérdezés, számlakivonatkérés vagy bankkártyaletiltás.

Egy másik komplexebb, de célzottan írott információn alapuló megoldás az Avatar, mely a vállalati hitelbírálat folyamatát kiegészítő gépi fórumtagként erősítette az akkori Fókusz Takarékszövetkezet nyolctagú hitelbíráló bizottságát. A működése során a csődelemzés folyamata egészül ki egy komplex szöveganalitikai modullal, amely mind a rövid, mind a hosszú szöveges tartalmak vizsgálatát jelenti. Továbbá képes írásbeli kommunikációt kezdeményezni a bíráló tagokkal úgy, hogy a hitelkérelem további kockázatos szempontjai kerüljenek előtérbe. Képes reagálni a fórumtagok felvetéseire, bár az első fejlesztési periódus felhasználói tapasztalatai alapján ez még korlátozottnak bizonyult. A problémát az jelentette, hogy a kritikus kérdésekre érkezett válaszok értelmezését követően az új válaszok rendszerbe történő beépítése mindenképpen igényelt fejlesztői aktivitást. Ezáltal a második periódusban,

[1] https://www.gartner.com/smarterwithgartner/how-to-use-citizen-data-scientists-to-maximize-your-da-strategy

az ún. „öntanuló” fázisban korlátozottan és felügyelten ugyan, de a fórumtagok megerősítését követően egy-egy új válasz az adott kérdésre beépíthetővé vált az adott rendszerbe informatikai közbeavatkozás nélkül. A feladat legnagyobb nehézsége az összehasonlításra kerülő magyar nyelvű szövegek témaspecifikussága és rövid terjedelme volt.
A Hanga megoldás jelenleg is aktív megvalósulása a vasúti menetrendi utastájékoztató rendszer komplex támogatását végző Elvira. Elvira képes egy összetett dialóguson végighaladva lefolytatni egy vagy akár több menetrendi lekérdezést is, és egy irányított párbeszéden belül dinamikusan tud reagálni. A rendszer több összefüggő komponensből áll össze, amely a beszédszintézis, beszédfelismerés és szöveganalitika szoros együttműködésén alapul.
A bemutatott megoldások azt hivatottak szemléltetni, hogy magyar nyelven is lehetséges olyan rendszerek felépítése és működtetése, amelyek hatékonyan tudják támogatni az adott szakterület munkáját és képesek olyan feladatok automatizált elvégzésére, amelyek kevesebb emberi beavatkozást igényelnek. Ezáltal a humán erőforrás bonyolultabb feladatokra összpontosítható.

Névelem felismerés vs. entitáskinyerés - eszközök és alkalmazási lehetőségek

Pancza Judit

Clementine, Budapest, Magyarország, jpancza@clementine.hu

Kulcsszavak: szöveganalitika, entitáskinyerés, névelem felismerés

A szöveges dokumentumok feldolgozása, azokból entitások és kapcsolatok kinyerése idő és erőforrás-igényes feladat, amely régóta komoly kihívások elé állítja a nyomozó hatóságok és minden olyan szervezet munkatársait, amelyek a nem strukturált adatokból szeretnének releváns információt kinyerni. A szövegbányászat éppen ezt a manuális munkát hivatott segíteni, a névelem felismerés mondhatni alap szövegbányászati feladat: segítségével kinyerhetők egy adott dokumentumhalmazban előforduló tulajdonnevek, például személynevek, helyek, szervezetek. Tehetjük ezt „hagyományosan" szótárak és szabályok segítségével, gépi tanulással vagy nagy nyelvi modellek segítségével.

Önmagában azonban az, hogy személyneveket, cég- és intézménynéveket, azonosítószámokat, telefonszámokat tudunk szövegbányászati módszerek segítségével azonosítani még nem elég a hatékony elemzéshez: meg kell határozni például, mitől lesz entitás egy név – vannak születési adatai, van címe? -, mikor tekintünk két entitást egyezőnek, és mindezt az információt megfelelő sémában el kell tárolni ahhoz, hogy például kapcsolatháló formájában vizualizálni tudjuk a szöveges dokumentumokban rejlő információt és összefüggéseket. A séma és vizualizációs lehetőségek szempontjából is fontos, hogy miből képzünk entitást, attribútumot és kapcsolatot: például egy cím lehet attribútuma egy személy entitásnak (mint lakcím), de önálló entitásként kell tárolnunk, ha szeretnénk olyan összefüggéseket is ábrázolni, mint a közös cím.

Az entitás azonosítási feladat tehát lényegesen túlmutat a klasszikus névelem felismerésen mind tervezési, mind feldolgozási szempontból. Hogy milyen módszert, eszközt használunk az adott feladathoz azt a cél, a munkamódszerünk és az adathalmazunk nagysága is meghatározza: nem mindegy például, hogy automatikus elemzést szeretnénk végezni tömeges dokumentumokon fix elemzési szempontok szerint vagy pedig a manuális kiértékelő, elemző munka támogatása a cél.

Az előadás során bemutatjuk azokat a szempontokat, amelyeket egy ilyen típusú feladat tervezésekor figyelembe kell venni, valamint azokat az eszközöket, amelyek hatékonyan képesek elvégezni a feladatot.