The background features a complex, abstract network of glowing blue lines and points, forming a series of interconnected triangles and polygons. The lines vary in opacity, creating a sense of depth and connectivity. The overall color palette is a range of blues, from deep navy to bright, glowing cyan.

**AZ OKTATÁS, A KUTATÁS ÉS
A KÖZGYŰJTEMÉNYEK DIGITÁLIS
TRANSZFORMÁCIÓJA FELSŐFOKON**

**NETWORKSHOP 2024
33. Országos Informatikai Konferencia**

**2024. április 3–5.
Eszterházy Károly Katolikus Egyetem, Eger**

AZ OKTATÁS, A KUTATÁS ÉS A KÖZGYŰJTEMÉNYEK DIGITÁLIS TRANSZFORMÁCIÓJA FELSŐFOKON

NETWORKSHOP 2024
33. Országos Informatikai Konferencia

2024. április 3–5.
Eszterházy Károly Katolikus Egyetem, Eger

Szerkesztette: Tick József, Kokas Károly, Holl András

HUNGARNET Egyesület
Budapest, 2024



HUN-REN
Magyar Kutatási Hálózat

NETWORKSHOP

Szerkesztette: Tick József, Kokas Károly, Holl András

Tipográfia és tördelés: Vas Viktória

Korrektúra: Danyi Melinda

Angol nyelvi lektor: Cseresnyés Dorottya

Networkshop 2024 konferencia előadásainak közleményei

Eszterházy Károly Egyetem, Eger

2024. április 3–5.

ISBN 978-615-82243-2-1

DOI: <https://doi.org/10.31915/NWS.2024>

Kiadja a HUNGARNET Egyesület
az MTA Könyvtár és Információs Központ közreműködésével

Budapest

2024

Borítókép: [freepik.com](https://www.freepik.com)

TARTALOMJEGYZÉK

Előszó	5
Ungváry Rudolf A MARC21 formátum kettős szerkezete és a formátum felhasználói szintjének fordításai	7
Holl András, Andódy Katalin Adatbányászati gyakorlatok repositóriumra és MTMT-re.....	16
Simon András Mesterséges intelligenciával támogatott adatgazdagítás a Nemzeti Levéltárban.....	22
Soós Gábor, Rövid András, Ormos Pál V2X – A járművek közötti kommunikáció kihívásai	29
Csernai Zoltán Egy online tanulást támogató portál kurzusának vizsgálata Big Data adatelemző módszerekkel.....	36
T. Nagy László, Németh Áron A mesterséges intelligencia (MI) teológiai kompetenciái	45
Mészáros Erika Kodaktól a jövőig – Egy könyvtári digitalizálás szinterei.....	55
Frankó Máté, Sándor Ákos Adatvizualizáció a könyvtári menedzsmentben: fejlesztések az SZTE Klebelsberg Könyvtár döntéstámogató rendszerében.....	63
Hernek István A felhasználóképzés szintjei az SZTE Klebelsberg Könyvtárban: az elsőévesektől a kutatókig	72
Némethi-Takács Margit, Borbély Mária Bibliográfiai kapcsolatok az általános megjegyzés adatmezőben.....	78
Dobás Kata, Tüskés Anna A magyar irodalomtörténet bibliográfiájának migrációja az ITIdata szemantikus adatbázisba	87
Horváth Péter A kanonikus magyar költészet versformakeresője.....	96
Sebestyén Ádám, Sárközi-Lindner Zsófia Történeti források szemantikus feldolgozása – Az ELTEdata adatbázis új gyűjteményei	105

Bolya Mátyás	
Lyukkártya és népdalrendezés – Egy mechanikus népzenei adatbázis digitális rekonstrukciójának lehetőségei.....	112
Kovácsházy Tamás	
Az idő, mint alapvető infrastruktúra, az idő szerepe az adatközpontban.....	121
Albert Ágota Katalin	
A mesterséges intelligencia használatának követelményei az oktatási szektorban, különös tekintettel a mesterséges intelligencia használatáról szóló rendeleltre.....	129
Varga Emese	
Digitális szövegszerkesztés a dHUpla keretrendszerében	135
Nemoda Zsuzsanna, Héjja Balázs, Nagy Andor, Tóth Máté	
A Pest Megyei Digitális Könyvtár fejlesztése	141
Nagy Dóra, Sándor Ákos	
Voice2text: a hanganyagátírás lehetőségei MI segítségével.....	149
Kalcsó Gyula	
Képek és metaadatok gyűjteményezése scrapingtechnológiával közösségi képmegosztó oldalról	157
Péter Róbert, Szántó Zsolt, Biacsi Zoltán, Kocsis Zoltán, Berend Gábor, Bilicki Vilmos	
Az AVOBMAT (Analysis and Visualization of Bibliographic Metadata and Texts) többnyelvű kutatási eszköz munkafolyamata és új funkciói	163
Máray Tamás	
Kvantum-számítástechnika: ez már a „jövő”?.....	171
Fellegi Zsófia	
Digitális kiadások migrációja: gépi és emberi intelligencia együttműködése.....	177
Palkó Gábor	
Posztmodern intertextualitás és digitális szövegkiadás	184
Antal Dániel	
A szlovák adatkicserélési tér magyarországi föderációjának lehetőségei.....	192
Vass Johanna	
Kutatási adatok megosztása a gyakorlatban – Adatrepozitóriumok használata az Ökológiai Kutatóközpont publikációiban	199
Mihály Eszter, Micsik András, Nagy Kadosa	
Irodalmi levélváltások nyomában TEI-vel és térképpel.....	208

Voice2text: a hanganyagátírás lehetőségei MI segítségével

Voice2text: the possibilities of voice transcription with AI

Nagy Dóra

Szegedi Tudományegyetem Klebelsberg Kuno Könyvtára

dora.nagy@ek.szte.hu

Sándor Ákos

*SZTE Informatikai és Szolgáltatási Igazgatóság,
Könyvtár-informatikai és Adatgazdálkodási Egység*

akos.sandor@ek.szte.hu

Absztrakt

Az MI-technológia fejlődésével manapság már rengeteg eszköz áll rendelkezésünkre hangfelvételek szerkeszthető szöveggé alakítására, amelyek mind sokkal egyszerűbbnek és gyorsabbnak tűnnek, mint a hallgatás utáni átgépelés. A korábban szinte csak angol nyelven elérhető programok többsége ma már boldogul a magyar nyelvvel is. Azonban kérdés, hogy ezek mennyire pontosak? Mennyi időt vesz igénybe, amíg kijavítjuk a hibákat? Kell-e minden esetben kézzel javítani vagy elfogadható a nyers gépi átírat? A cikkben röviden bemutatásra kerülnek különböző beszéd felismerő programok, illetve egy konkrét projekten keresztül ismertetjük saját tapasztalatainkat is.

Kulcsszavak: MI, mesterséges intelligencia, beszéd felismerés, hanganyagátírás

Abstract

In this article, different methods of transcribing audio material are presented through concrete examples. With the advances in AI technology, there are now many tools available for converting audio recordings into editable text, all of which seem much easier and faster than transcribing after listening. Many of the programmes that were previously available almost exclusively in English now also work in Hungarian. But how accurate are they? How long does it take to correct errors? Do they always have to be corrected by hand, or is the raw machine transcription acceptable? In this article, we will introduce different speech recognition software and present our own experience through a specific project.

Keywords: AI, artificial intelligence, speech recognition, voice transcription

Bevezetés

A téma apropóját az SZTE Móra Ferenc Szakkollégiumban több évtizedes múltra visszatekintő előadások felvételeinek SZTE Klebelsberg Könyvtárba kerülése adta. A felvételek szolgáltatása mellett felvetődött a kérdés, hogyan lehetne még szélesebb körben hozzáférhetővé tenni a hanganyagokat? Ha készülné hozzájuk átírat, azzal egyrészt indexelésre kerülhetnének a teljes szövegű (full-text) keresőben, másrészt a javított átírat feltöltése a hangfájl mellé lehetőséget biztosítana a hallássérült felhasználóknak a tartalom megismerésére. Emellett a több forrásból származó hang- és videóanyagokhoz készített átírat tovább növeli a szolgáltatás színvonalát, így érdemes élni ezzel a lehetőséggel. Ilyen anya-

gokat jelentenek például gyűjteményünkben a Szent-Györgyi Albert hagyatékban található orsós magnószalagok és hangkazetták. További nagy egységet jelent az SZTE Elektronikus Tananyag Archívum oktatók által feltöltött több ezer videóleckéje és hanganyagai.

Móra Digitális Hangarchívum

Az SZTE Móra Ferenc Kollégiumban az 1980-as évek elejétől kezdve napjainkig rendeznek különféle előadásokat, író-olvasó találkozókat, beszélgetéseket, emlékesteket, fórumokat, interjúkat, folyóirat- és irodalmi esteket. A 2010-es évek elejéig rögzítették ezeket magnókazettára, majd digitalizálták CD-lemezekre.

Az előadások rendkívül változatos, irodalmi, társadalomtudományi és filozófiai témákat járnak körül. A résztvevők, előadók között olyan jeles személyiségek, irodalmárok és történészek szerepeltek, mint Csoóri Sándor, Esterházy Péter, Faludy György, Heller Ágnes, Illia Mihály, Kertész Imre, Kosáry Domokos, Krasznahorkai László, Nádas Péter, Parti Nagy Lajos és Vásárhelyi Miklós.¹

Az SZTE Klebelsberg Könyvtár 2022 novemberében kapta meg a CD-re mentett változatokat, ami összesen 439 darab lemezt jelentett. A CD-k mentésekor derült ki, hogy az előadások pár perces darabokban vannak a lemezeken, így első lépésként megfelelő sorrendben össze kellett illeszteni a fájlokat. Az összefűzésére az FFmpeg² programcsomagot használtuk.

Mivel ezek régi felvételek és többszörös konvertáláson estek át, sok esetben a hang minősége nem tökéletes. Némelyik felvétel nagyon zajos és előfordul, hogy másodpercek hiányoznak belőle. Zajcsökkentésre és hangjavításra az Audacity³ szabad és nyílt forráskódú több platformos digitális audió-szerkesztő programot használtuk. Kipróbáltuk a beszédminőség javítására az Adobe Enhanced Speech⁴ online mesterséges intelligencia szoftvert is. Az ingyenes verzióban napi 1 órányi hanganyagot lehet feltölteni. A zajszűrés nagyon jól sikerült, viszont a beszédet is próbálta javítani, ami a magyar nyelvű anyagok esetében sajnos nem működött. Jelenleg nem lehet nyelvet választani, így mindent angol nyelvűként értelmez és aszerint próbálja kijavítani a szöveget. Mivel túlnyomó részben magyar nyelvű anyagaink vannak, ezt a szoftvert egyelőre elvetettük, de későbbiekben hasznos lehet az angol nyelvű felvételek esetén. Készült egy Python script is, ami FFmpeg² és bandpass-filtert használva próbálja automatikusan felismerni és javítani a zajos részeket. Az automatikus átírat ugyanis gyakran „AAAAA” karaktersorozatot tesz az átíratba a zaj átírásaként.

Összesen több mint kétszáz előadás anyagát kaptuk meg, melyből jelenleg 110 elérhető az SZTE Klebelsberg Könyvtár Képtár és Médiatékájában⁵. A metaadatok egyrészt a CD-k borítójáról, másrészt a kollégium saját nyilvántartásából álltak rendelkezésre. Készítettünk

1 Panek Sándor (2024): Egy ládányi szabad beszéd – A Móra Kollégium legendás találkozói az SZTE Klebelsberg Könyvtár Médiatékájában. *SZTE hírek*, 2024. május 22. <https://u-szeged.hu/sztechirek/2024-majus/mora-kollegium-talalkozok-hanganyaga-szte-klebelsberg-konyvtar> (Utolsó letöltés: 2024.06.18.)

2 FFmpeg hivatalos oldala: <https://ffmpeg.org/>

3 Audacity hivatalos oldala: <https://www.audacityteam.org/>

4 Adobe Enhanced Speech hivatalos oldala: <https://podcast.adobe.com/enhance#>

5 Móra Ferenc Kollégium szakmai előadásai, kurzusai: <https://mediateka.ek.szte.hu/collections/show/89>

ezek alapján egy táblázatot, mely lehetővé tette a tömeges betöltést is. Összesen körülbelül 400 órányi hanganyagra becsüljük az egész gyűjteményt. Sajnos vannak hiányzó CD-k, így néhány előadás nem teljes. Szerencsére a kollégiumban megvannak az eredeti kazetták, így van esély arra, hogy előkerülnek még hiányzó darabok is.

Célok

Mielőtt elkezdtük a tényleges átíratok készítését, két célt fogalmaztunk meg: az egyik szerint csak automata átírást használunk, melynek lényege a full-text kereshetőség megvalósítása. Gyakorlatban ez annyit jelent, hogy a könyvtár digitalizált tartalmainak közös kereső indexében (Contentas⁶) rejtett szövegfájlként tároljuk a megfelelő rekordhoz kapcsolva a hanganyagok teljes szövegét, így a karakterfelismert PDF-dokumentumainkhoz hasonlóan ezekre is működik a teljes szövegű keresés.

Ezen kívül felmerült egy másik lehetőség is, miszerint a korrektúrázott átírat PDF-formátumban bekerülhet a rekordba a hangfájl mellé, így olvasható és kereshető is lesz a tartalom. A javított szövegátírat hozzáférést biztosít a hallássérült felhasználóknak a hanganyagokhoz. Emellett könnyebben idézhetővé és felhasználhatóvá válnak ezáltal az anyagok. Az automata átírat nem kerülhet ki ellenőrzés és korrektúra nélkül a rekordba, mivel az OCR-hez hasonló hibák a hangfelismerésnél is előfordulnak. Viszont amíg a szöveges dokumentumokból keletkezett kétrétegű PDF esetében az olvasó mindig az eredeti nyomtatott szöveget látja, vagyis nem annyira szembetűnő a hiba, addig a hanganyagoknál nincs ami elfedje ezeket.



1. ábra: Automata átírat vs. javított szöveg

Hangátírás

Több online elérhető hangfelismerő programot próbáltuk ki, melyek a következők voltak: Alrite, Transkriptor, Microsoft Office 365 Word, Microsoft Azure Speech to text API.

Alrite⁷

Az Alrite gépi tanulásra épülő beszédfelismerő szolgáltatás. Fejlesztője a magyar Régens Zrt., több mint 20 éves szakmai tapasztalattal rendelkező informatikai vállalat. Profiljuk az üzleti szoftverfejlesztés, felhő- és IT-szolgáltatások. Az Alrite 2020 eleje óta érhető el, azóta is folyamatosan fejlesztik. Kifejezetten magyar nyelvre optimalizált, de összesen 12 nyelvet ismer: angol, dán, francia, holland, lengyel, magyar, német, olasz, portugál, spanyol, svéd, török. Az általánosan ismert audio és videó fájlkiterjesztéseket támogatja (pl. WAV, MP3, MP4).

6 SZTE Klebelsberg Könyvtár Repozitóriumok közös keresője <https://contentas.bibl.u-szeged.hu/>

7 Alrite hivatalos oldala: <https://alrite.io/ai/hu/>

A felületen működik a diktálás, fájlfeltöltés saját gépről, online link beszúrás, élő közvetítés feliratozása is. Készíthető vele időbélyeges felirat és sima folyószöveges átírat. Online szövegszerkesztőben lehet javítani a felismert szöveget, aminek része a bizonytalan szavak kiemelése, vagyis pirossal megjelöli a felismeréskor a rendszer szerint bizonytalannak ítélt kifejezéseket. A beszélők megkülönböztetése is automatikusan történik. Egyenként maximum 1 GB méretű fájl lehet feltölteni. Használható böngészőből és mobil alkalmazásból is. Saját bevallásuk szerint 90–95% pontossággal dolgozik a rendszer⁸, a teszt feltöltések során mi is ezt tapasztaltuk.

Több verzió érhető el, melyek között van ingyenes és fizetős is. A Starter verzió esetében egy természetes személy csak egy e-mail címen jogosult ingyenes tesztverzió hozzáférést igényelni. Havi 30 perces időkeret áll rendelkezésre és a regisztrációt követően 5 évig érhető el. A rendszer hét napig tárolja a fájlokat és az átíratot, utána törlésre kerül minden, ezért figyelni kell, hogy időben lementésre kerüljenek az anyagok. A videófeliratozás a funkció megismerésének és tesztelésének céljából csak az első 2 percre készül el. Fordítás nem érhető el az ingyenes verzióban.

Az előfizetési verziókban minden funkció elérhető. A díjcsomagokra előfizetőknek helpdesk-szolgáltatást is biztosítanak. A szolgáltatás része az automatikus fordítás és az időbélyegeket tartalmazó leirat készítése. Az adminisztrációs felület korlátlan felhasználó kezelését biztosítja, valamint a fájlokat és szövegváltozatokat egy évig tárolja. Kiegészítő üzleti szolgáltatásokat is biztosítanak, melynek része az API-elérés, tömeges fájlfeldolgozás, valós idejű beszédfelismerés, privát felhő, egyedi kifejezésekre tanítás, egyedi fejlesztések.

A felület jól átlátható, használata egyszerű. Az átíratot gyorsan elkészíti: egy 23 perces anyagnál körülbelül 2–3 perc alatt, melyet a felületen lehet javítani, illetve letölteni.

Transkriptor⁹

A következő, amit kipróbáltunk az amerikai Transkriptor volt. Több, mint 140 nyelven képes átíratokat készíteni, köztük magyarul is. Nagyon sok audio és videó fájlkiterjesztést támogat (pl. MP3, MP4, WAV stb.). Ennél a programnál is működik a diktálás, fájlfeltöltés saját gépről, online link beszúrás és élő közvetítés (online meeting). Képes generálni időbélyeges felíratot és egyben folyószöveges leíratot is, illetve a beszélők megkülönböztetése automatikusan történik. Használható böngészőben, mobil alkalmazással és készíttetek hozzá Google Chrome bővítményt is. Zoom, Microsoft Teams és Google Meet rendszerekbe is integrálható. Az ingyenes verzióban maximum 1 GB méretű fájlokat lehet feltölteni. Saját bevallásuk szerint 99% pontossággal dolgozik a rendszer¹⁰, mely a tesztek alapján a magyar nyelv esetében kicsit kisebbnek tűnik.

Az ingyenes próbaverzió keretében egyszeri 90 percet kap egy felhasználó, viszont az átírat minden esetben a feltöltött fájl első 5 percét készíti el. A beépített Google fordító segítségével azonnal készíthető hozzá fordítás. Kiegészíti egy online szövegszerkesztő, ahol javítani lehet az automatikus átíratot, ill. módosítani a beszélőket. Viszont az időbélyeget nem

8 Alrite felhasználási feltételek: <https://alrite.io/ai/hu/terms-of-use/>

9 Transkriptor hivatalos oldala: <https://transkriptor.com/hu/>

10 Transkriptor Hangból szöveg: <https://transkriptor.com/hu/audio-szoveg/>

lehet szerkeszteni és nem lehet se letölteni, se kimásolni a szöveget. Előfizetés nélkül csak arra jó, hogy kipróbáljuk, lássuk a felületet, de egy komplett átírást nem lehet az ingyenes verzióval elkészíteni.

Microsoft Office 365 Word Web App¹¹

A harmadik program, amit végül a kipróbáláson kívül a gyakorlatban is rendszeresen használtunk a Microsoft Office 365 Word Web App. A Szegedi Tudományegyetem rendelkezik Microsoft 365 előfizetéssel, így szinte korlátok nélkül tudtuk használni. Összesen 82 nyelvet ismer, köztük a magyart is. Nem támogat annyira sokféle fájlformátumot, mint például a Transkriptor, csak néhány általánosan ismert fájlkiterjesztés használható: WAV, MP4, M4A, MP3. Időbélyegek és beszélők megkülönböztetése ebben az esetben is automatikusan megtörténik. Elérhető a diktálás, fájlfeltöltés saját gépről, viszont online tartalmat linkként beillesztve nem tud átírni. Javasolt a Microsoft Edge vagy Chrome böngésző használata. Egy felhasználó havonta 300 percnyi hanganyagot tud átírni vele. Az átírat és a hangfelvétel tárolása OneDrive tárhelyen történik, ahol megosztás is lehetséges, illetve letölthetők saját gépre is. Fájl méret korlátozás így gyakorlatilag nincs, a OneDrive tárhely méretétől függ, hogy mennyi anyagot tud tárolni. Az átírat szerkesztőben lehet átcímkezni az előadót, javítani a szöveget, hozzáadni és törölni időbélyeget, illetve a lejátszás sebessége is módosítható. Az átírt szöveg beillesztésének lehetőségei a dokumentumba a következők: csak szöveg, hangszórókkal (melynek segítségével az adott résztől indul a hanganyag lejátszása), időbélyegekkel, hangszórókkal és időbélyegekkel. A munkafolyamat elkezdéshez egy új, üres dokumentumot kell megnyitni, majd a menüsávban az átírás gombra kattintva először ki kell választani a nyelvet, majd betallózni a hangfájlt. Ezután elkezdődik a feltöltés a OneDrive tárhelyre, majd az átírás¹². Egy órás hanganyag esetében is csak 2–3 perc az egész folyamat (feltöltéssel és automata átírással). A tapasztalatok azt mutatják, hogy nagyon kényelmesen és jól lehet vele dolgozni. Az átírat javítására fordított munkaidő átlagosan 4-6 munkanap, ami természetesen függ az előadás témájától és a hanganyag hosszától is. Előfordult, hogy egy-egy eseménynek, vagy személynévnek utána kell nézni, mert hallás után nem egyértelmű miről is van szó.

Microsoft Azure Speech to text API¹³

Amilyen gyors az automata felismerés olyan lassú a kézi javítás, ezért ezt a módszert nem tudjuk és nem is szeretnénk minden hanganyag esetében alkalmazni. A másik célunk megvalósítására (vagyis az automata átírás indexbe rejtett text-fájlként való használata a full-text kereshetőség biztosítása érdekében) a Microsoft Azure Speech to text API-t használtuk.

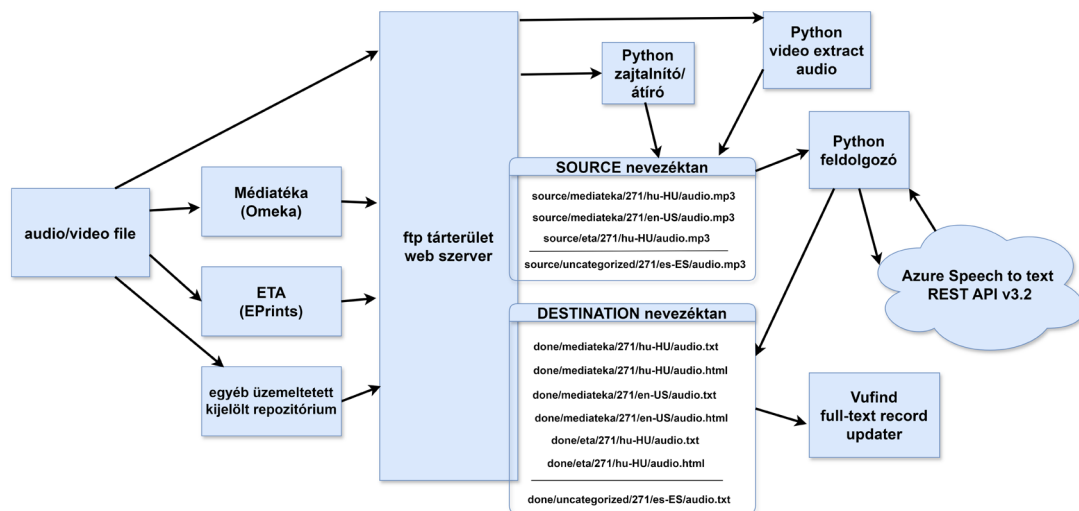
Használható a hangstreamek valós idejű vagy kötegelt átírására szöveggé (batch-átírás). Mesterséges intelligencia működik a háttérben és mindent, amit az Office 365-ön keresztül

11 Microsoft Office 365 Word Web App hivatalos oldala: <https://www.microsoft.com/hu-hu/microsoft-365/word>

12 A felvételek átírása Webes Wordben: <https://support.microsoft.com/hu-hu/office/a-felv%C3%Agtel%C3%Ait%C3%ADr%C3%A1sa-7fc2efec-245e-45fo-b053-2a97531ecf57?ui=hu-hu&rs=hu-hu&ad=hu>

13 Microsoft Azure Speech to text API-ismertető a hivatalos oldalon: <https://learn.microsoft.com/hu-hu/azure/ai-services/speech-service/speech-to-text>

el lehetett érni az az API-n keresztül is használható. A kötegelt feldolgozásnak köszönhetően nem kell egyesével feltölteni az átírandó anyagokat. REST API használatával történik az átírás, melyből a v3.2 preview API-verziót használjuk (v3.1 a default verzió jelenleg). Többféle belső beszédmodell érhető el, sőt saját egyedi is tanítható (persze ott már az árképzés is módosul). Konfigurálásnál az API-nak nem kötelező modellt megadni, ilyen esetben az alapértelmezettet használja. Az árképzés a következőképpen alakul pillanatnyilag: €0,332/óra → 100 óra €33.



2. ábra: A munkafolyamat sematikus rajza

A fenti munkafolyamat-ábrán látszik, hogy az audio- és videófájlok egyrészt valamelyik repozitóriumból érkeznek, vagy valamilyen más rendszerből, illetve még nem repozitált anyagok is lehetnek.

A munkafolyamat első lépése a céljainknak megfelelő FTP-tárterület, webszerver létrehozása volt. Ahol helyi szerver nem elérhető, abban az esetben az Azure felületén is van lehetőség Blob Storage létrehozására az audio- és videófájlok tárolására, azonban ez pluszköltséget jelent.

Következő lépés az Azure API-kulcs generálása a speech2text szolgáltatáshoz, amit a Python feldolgozó script használ az automatikus átírás során. Ezután az audio- és videófájl átkerül egy helyi FTP-szerverre, ahonnan a 'source'-mappából egy Python script feldolgozza és a 'destination' mappába leteszi az automatikus átíratot. A Python script főbb feladatai: zajszűrés¹⁴; audio hangszáv kiszedése a videófájlból¹⁵; Azure API-kommunikáció.

API-beállításoknál az alábbi adatokat kell megadni: Content URL, Project name, Locale (azaz az audio nyelve, pl. hu-HU), DiarizationEnabled (beszélő felismerése: Voice1, Voice2 stb.), Channels (0, azaz csak a bal oldali csatorna), Regio, API version. Kimeneti formátum-

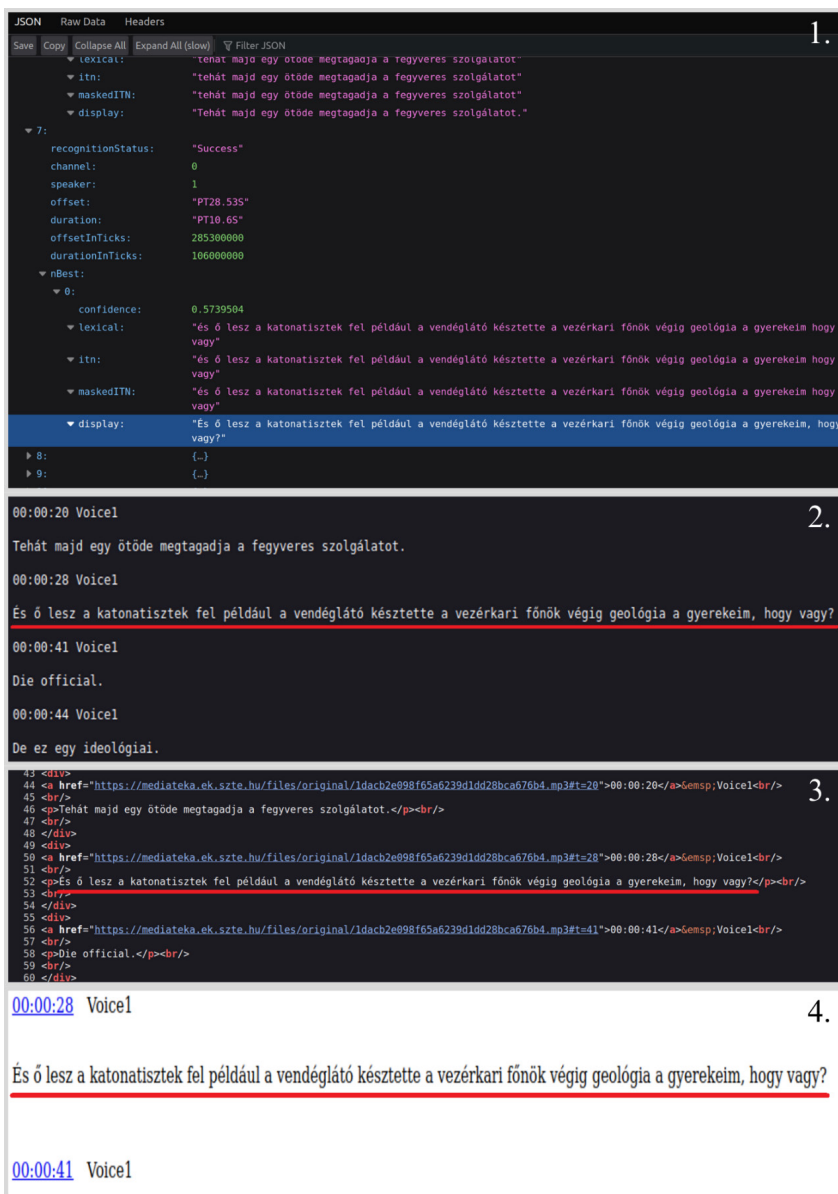
14 ffmpeg -i AUDIOFILE outputpath -loglevel quiet -y -af highpass=f=200, lowpass=f=3000

15 ffmpeg -i VIDEOFILE outputpath -loglevel quiet -y

ként JSON-t használ, melyben érdekes adat a 'confidence'-érték, mely azt mutatja mennyire biztos a felismerés pontosságában.

A speech2text JSON-exportból a Python alkalmazás elkészít egy olyan TXT-fájlt, amelyben már csak a full-text indexeléshez szükséges szöveg szerepel. Ez egy Vufind-specifikus JSON-be töltődik át, amely már közvetlenül betölthető a Solr-be.

A metaadatok közül az adatforrás neve (pl. ETA, Médiatéka) és a rekord ID szerepel az átalakítandó TXT-fájl elérési útvonalában és ennek segítségével könnyen lehet frissíteni a Solr-indexet az automata szöveggel.



3. ábra: API-munkafolyamat állomásai

Abban az esetben, ha szükséges a hanganyagok kézi javítása, akkor az elkészült PDF beke-
rül a repozitóriumba majd az automata szövegátírat felülíródik a rekord újraindexelésekor.

A fenti első képernyőfotón az látható, hogy az API-n keresztül visszakapott átíratok közül a 'display' értékét használjuk fel a céljainkra. A második képernyőfotón egy technikailag köztes txt látható. A harmadik képernyőfotó annak a HTML-fájlnak a forráskódja, melyet a Python script készít (melynek segítségével a kézi javítások készülnek). A negyedik képernyőfotó a böngészőbeli nézete ennek a HTML-fájlnak a kattintható időbélyeggel.

	Alrite	Transkriptor	Microsoft Office 365 Word	Microsoft Azure Speech to text v3.2 API
ingyenes próba	havi 30 perc	90 perc	300 perc	300 perc
árak, intézményi felhasználás	53, 80, egyedi Ft / perc	\$30, egyedi / hó	€4.60 felhasználó / hó	€0,332 / óra
árak, magán célú felhasználás	33, 44 Ft / perc	\$4.99, \$12.49 / hó	3599 Ft / hó	-
nyelv	12 (magyar is)	143 (magyar is)	82 (magyar is)	82 (magyar is)
online szövegszerkesztő	igen	igen	igen	nem
beszélők megkülönböztetése	igen	igen	igen	igen
átírat	teljes	5 perc	teljes	teljes
támogatott kiterjesztések	WAV, MP3, M4A, OGG, MP4, WEBM, MOV, Opus	MP3, MP4, WAV, AAC, M4A, WebM, FLAC, Opus, AVI, M4V, MPEG, MOV, OGV, MPG, WMV, OGM, OGG, AU, WMA, AIFF, OGA	WAV, MP4, M4A, MP3	WAV, MP4, M4A, MP3
átírat fájlformátum	SRT, Word	TXT, SRT, Word	Word	JSON

4. ábra: Hangfelismerő programok összehasonlítása (készült 2024. április)

A gyakorlatban végül a Microsoft alkalmazások használata mellett döntöttünk, melynek egyrészt praktikus oka van: a Szegedi Tudományegyetemnek ezekre a programokra van előfizetése. Másrészt a használat során szerzett pozitív tapasztalatok miatt is ezeket választottunk volna. A Microsoft Office 365 Word Web App a klasszikus grafikus felületnek köszönhetően egyszerűen kezelhető, emellett gyorsan és nagy pontossággal ismeri fel a hanganyagokat. A Microsoft Azure Speech to text API konfigurálásához informatikus szak- tudás szükséges, de a befektetett munka megtérül a tömeges fájlfeldolgozás megvalósításá- val. Így szívesen tudjuk ajánlani mindkét rendszert bármely közgyűjteménynek.

A további fejlesztési tervek, lehetőségek között szerepel a videók feliratozása, mely szintén megoldható Microsoft Azure Speech to text API-val. Illetve automatikus fordítás elkészí- tése, melyhez Azure Cognitive Services - Translator Text API-ra van szükség.

Közreműködő kollégák: Bocsán János Gergő, Csengődi Erika, Dávid Norbert, Erdődi Zol- tán, Farkas Richárd, Fekete Franciska, Házi Csaba, Kiss Fatima Jázmin, Nagy Gyula, Szalai Fruzsina, Zsemberi Adrienn.

A fejlesztés megvalósulását az RRF-2.1.2-21-2022-00012 azonosítószámú, Komplex Digitális Modellváltás – intelligens fokozatváltás projekt támogatja. A projekt Magyarország Hely- reállítási és Ellenállóképességi Tervének keretében valósul meg.