

BESZÉDFELISMERÉS MAGYARUL

Szavakból is értenek



Ami akár még öt éve is vágyalomnak tűnhetett, mára valósággá vált: több általános célú, magyarul (is) tudó beszédfelismerő szoftver közül válogathat, aki nem akar többé leleteket, határozatokat, testületi ülések jegyzőkönyvét vagy éppen interjúkat begépelni.

Természetesen mindenkit az érdekel a legjobban, hogy mennyire pontosan ismerik fel ezek a rendszerek az élőbeszédet. Erre azonban gyakorlatilag lehetetlen egzakt választ adni. A legjobb algoritmus is rosszabbul teljesít, ha túl zajos a felvétel, míg fordítva, egy kiváló mikrofon, az alacsony háttérzaj és az ismert szakmai szövegek környezetét gyakorlatilag 100 százalékos pontosságot ígér. Az alább bemutatott megoldások mindegyike jól teljesít a saját területén – azt kell eldönteni, hogy mi is azon a területen mozgunk-e.

Radiológusul is tud

Az orvosi szakirányú felhasználás irányából érkezett az általános beszédfelismerés világába a Belux Csoport, és terméke, a Globalspeech szoftver. „2005-ben fejlesztettük ki az első szoftvert radiológusok számára leletek diktálására”, mesél a kezdetekről *Benes Edvárd*, a vállalat cégvezetője. Az első verziót később műtéti, patológiai, majd jogi beszédleíró követte, míg tavaly elkészült az általános célra használható megoldás is. A Globalspeech a mesterséges intelligenciára építve folyamatosan tanul, fejlődik, megszokja a felhasználó beszédstílusát, ritmusát, külön szótanító funkciót alkalmaz. A szakmai szótáraknak, speciális nyelvtani moduloknak és algoritmusoknak köszönhető a közel százszázalékos felismerési arány, s hogy a szoftver például diktáláskor nem azt írja le „a bétékát tizenegyedik paragrafus első bekezdése szerint”, hanem „A Btk. 11. § (1) bekezdése szerint...”. Különlegessége még a Globalspeech-nek, hogy 17 nyelven érhető el, jól ismert Szerbiában, Görögországban és Cipruson is. Közvetlenül a Wordbe lehet vele diktálni, de leírhatók vele korábban felvett hanganyagok és videók is.

A Globalspeech szoftver előnye, hogy különböző információs rendszerekbe integrálható (Wordbe, LibreOffice-ba stb.) A felhasználó gépén, hálózati vagy internetelérés nélkül is működik. A felismerés és a szövegek a cég vagy az ügyfél szerverein tárolódnak. Különösen érzékeny – jogi, egészségügyi, céges – adatok esetekben akár a kliensgépen, cache-ből történhet a felismerés. Elkészült a webes és a mobil verzió is. Hazai szerverparkból üzemelő szolgáltatást is nyújt a társaság.

Az egyszerűség ereje

Szintén saját algoritmust használ Alrite nevű megoldásában a Régens. A vállalat a mesterséges intelligencia egyéb (arcfelismerés, objektum- és szövegelemzés) területein is rendelkezik tapasztalattal, mondta el érdeklődésünkre *Matics-Horváth Kata*, a vállalat marketingigazgatója. Az Alrite rendkívül egyszerűen használható. Web alapú (és mobilalkalmazásként is elérhető) rendszer, amelyhez csak egy rövid regisztráció kell, és megújuló havi időkorláttal ingyenesen is használható – bizonyos

Okosan fordít

Kimondottan fordítók számára készült az e-dictate szoftver. A webes felületen is elérhető megoldás szinte valamennyi funkciója arra szolgál, hogy megkönnyítse a fordítók munkáját. Nemcsak tanítható saját kifejezésekkel, de hangparancsokkal is kezelhető. Diktálhatók neki formátumcímkek, írásjelek, idéző- és zárójelek, felismeri és helyesen írja le a mértékegységeket, számokat, és áttemelhetők a fordításba kifejezések a forrásszövegből.

korlátokkal. Ezért jó választás azoknak, akik nagyobb infrastruktúra-beruházás nélkül akarnak leiratokat készíteni.

Négyféle funkciót kínál: feltöltött hang- és videóállományok vagy diktálás leírása a magyar mellett angol és német nyelven, illetve live stream és YouTube-videók feliratozása. A leírt szöveg az Alrite saját ablakában javítható, ahol a kérdéses szövegrészletre ugorva egyből meghallgatható a felvétel is. A leirat fordítható, és több formátumban letölthető.

Nem ír elő ajánlott mikrofont vagy hangrögzítőt az Alrite, de tapasztalataink szerint viszonylag zajos felvétellel is jól boldogul. A szoftver több beszélőt is képes megkülönböztetni, és ennek megfelelően tagolni a szöveget, mint ahogy jól megbirkózik a központozással is. A hangállományokat és a szövegeket saját, Magyarországon működő szervereiken tárolják, de azokat nem használják az MI tanítására, az algoritmust más forrásokból fejlesztik.

A webes felhasználás mellett itt is van lehetőség saját szerverekre történő telepítést kérni (a szoftver integrálható üzleti alkalmazásokkal is). „A Régens vállalja, hogy kérésre betanítja a rendszernek az ügyfél egyedi szakkifejezéseit”, tette hozzá *Matics-Horváth Kata*. A továbbfejlesztési tervek között szerepel egyebek mellett videokonferenciák valós idejű feliratozása, az érzelmek detektálása, és további nyelveken történő leiratozás és fordítás.

Az alacsony zajszint és az ismert szakmai szövegek környezet közel százszázalékos pontosságot ígér: 100-ból 5-nél kevesebb szót kell javítani a leiratban

A mindenes

Alapesetben a Google beszédfelismerőjét használja a Gallmed Absonic nevű terméke, amely a felhasználók gépein fut, de igény esetén saját szerveres, offline megoldás is megvalósítható. Ebben az esetben egyedileg tanítják be a megoldást a felhasználó terminológiájára, és az Absonic saját algoritmusát használják a hangfelvételek leírásához, mondja *Várhalmi Antal Miklós*, a mikrovállalkozás ügyvezető igazgatója.

Az Absonic rendkívül széles körű felhasználhatóságával tűnik ki a mezőnyből. A diktálás, a rögzített felvételek és a tárgyalások leiratozása mellett a legkülönbözőbb távközlési csatornákon folytatott beszélgetések megjelenítésére és leírására is lehetőség van, a PSTN-től kezdve a GSM-en és az URH-n át a WhatsApp-ig; mindehhez a szükséges hardveres eszközöket is kínálja a vállalkozás. Alkalmas az Absonic videók (akár Youtube- és Facebook-videók) valós idejű feliratozására is. Különlegesség például, hogy prezentációkat is lehet vele valós időben feliratozni, és ezzel együtt egy másik nyelvű fordítást is megjeleníteni. A hallássérülteket támogatja az előadói, oktatói beszédek valós idejű megjelenítése, a kézhasználatban korlátozottaknak pedig módot ad a számítógép beszéd alapú vezérlésére.

A sokrétű funkcionalitás biztosításához az Absonic a felhasználók gépére telepítendő, mintegy 900 ezer soros programból és (offline igények esetén) szerveroldali komponensekből áll. A minél pontosabb felismerés érdekében érdemes biztosítani a megfelelő beszédminőséget zajszűréssel, automatikus szintkiegyenlítéssel, és célszerű professzionális mikrofont vagy beszédforrást használni. „Ha nem törődünk a bemeneti jel minőségével, a felismerés sem lesz elegendő pontosságú”, mondja *Várhalmi Antal Miklós*.

Schopp Attila